Data-driven Channel Learning for Next-generation Communication Systems

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Donghoon Lee

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Prof. Georgios B. Giannakis, Advisor

October, 2019

Acknowledgments

First and foremost, my deepest gratitude goes to my advisor Prof. Georgios B. Giannakis. I would like to thank him for giving me a great opportunity to be a member of the SPiNCOM research group, as well as a student of the University of Minnesota, Twin Cities. Without his guidance and insightful suggestions, this thesis would never been completed.

I would like to extend my appreciation to collaborators during my Ph.D. years, Prof. Seung-Jun Kim and Prof. Daniel Romero. Their vast knowledge and insight helped lay a cornerstone for this thesis. Special thanks go to Prof. Mostafa Kaveh, Prof. Mehmet Akçakaya, and Prof. Rui Kuang who agreed to serve as my thesis committee. Thanks also go to a number of other professors in the departments of Electrical Engineering and Computer Science, whose graduate level courses provided me with the necessary background to embark on this area of research.

I also would like to credit current and former members of the SPiNCOM research group for all the good times in Minnesota: Prof. Emiliano Dall'Anese, Prof. Gonzalo Mateos, Prof. Juan Andrés Bazerque, Prof. Antonio G. Marqués, Prof. Vassilis Kekatos, Prof. Yu Zhang, Dr. Morteza Mardani, Prof. Nasim Soltani, Dr. Brian Baingana, Dr. Dimitris Berberidis, Dr. Fatemeh Sheikholeslami, Dr. Panagiotis Traganitis, Dr. Gang Wang, Prof. Tiany chen, Prof. Yanning Shen, Dr. Liang Zhang, Prof. Jia Chen, Dr. Athanasios Nikolakopoulos, Dr. Qin Lu, Alireza Sadeghi, Georgios Karanikolas, Meng Ma, Vassilis Ioannidis, Bing Cong, Elena Ceci, and Seth Barrash; as well as my friends and colleagues from the Digital Technology Center and the University of Minnesota: Charilaos Kanatsoulis, Agoritsa Polyzou, Ioanna Polyzou, Andreas Katis, Maria Kalantzi, Dr. Vasileios Kalantzis, Dr. Dimitrios Zermas, Dr. Panagiotis Stanitsas, Nikolaos Stefas, Nikos Kargas, Dr. Mohit Sharma, Dr. Shaden Smith,

Dr. Evangelia Christakopoulou, Dr. Konstantina Christakopoulou, Dr. Aritra Konar, Dr. Bo Yang, Dr. Ahmed Zamzam, Faisal Almutairi, Ancy Tom, Dr. Daehan Yoo, Dr. Sehyun Hwang, Prof. Daeha Joung, Prof. Song Min Kim, Dr. Hyungjin Choi, Dr. Chanjoon Lee, Gyubaek Choi, Seonmo Kim, Jeehwan Song, Minsu Kim, Gyusung Park, Taejoon Byun, and Jungseok Hong.

Last but not the least, I am very grateful to my family: my parents Jin Seung Lee and Hawja Yoon, my parents-in-law Sung Won Kim and Kyeong Ok Cha, my grand parents Hee Duk Lee and Junghee Kang, my dear brother Juwon Lee, for their love and support throughout my life. Su, my one and only, I would never make this without you. Thank you for always being next to me.

Donghoon Lee, Minneapolis, October, 2019.

Dedication

This dissertation is dedicated to Su and Lana for their unconditional love and support.

Abstract

The turn of the decade has trademarked the 'global society' as an information society, where the creation, distribution, integration, and manipulation of information have significant political, economic, technological, academic, and cultural implications. Its main drivers are digital information and communication technologies, which have resulted in a "data deluge", as the number of smart and Internet-capable devices increases rapidly. Unfortunately, establishing information infrastructure to collect data becomes more challenging particularly as communication networks for those devices become larger, denser, and more heterogeneous to meet the quality-of-service (QoS) for the users. Furthermore, scarcity in spectral resources due to an increased demand for mobile devices urges the development of a new methodology for wireless communications possibly facing unprecedented constraints both on hardware and software. At the same time, recent advances in machine learning tools enable statistical inference with efficiency as well as scalability in par with the volume and dimensionality of the data. These considerations justify the pressing need for machine learning tools that are amenable to new hardware and software constraints, and can scale with the size of networks, to facilitate the advanced operation of next-generation communication systems.

The present thesis is centered on analytical and algorithmic foundations enabling statistical inference of critical information under practical hardware/software constraints to design and operate wireless communication networks. The vision is to establish a unified and comprehensive framework based on state-of-the-art data-driven learning and Bayesian inference tools to learn the channel-state information that is accurate yet efficient and non-demanding in terms of resources. The central goal is to theoretically, algorithmically, and experimentally demonstrate how valuable insights from data-driven learning can lead to solutions that markedly advance the state-of-the-art performance on inference of channel-state information.

To this end, the present thesis investigates two main research thrusts: i) channel-gain cartography leveraging low-rank and sparsity; and ii) Bayesian approaches to channel-gain cartography for spatially heterogeneous environment. The aforementioned research

thrusts introduce novel algorithms that aim to tackle the issues of next-generation communication networks. Potential of the proposed algorithms is showcased by rigorous theoretical results and extensive numerical tests.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Context

Smart and Internet-capable devices have a ubiquitous presence in our daily lives. In the first quarter of 2019, global mobile penetration is 104 percent, bringing the total number of mobile subscriptions to around 7.9 billion. Correspondingly, the global mobile data traffic grew by 82 percent between 2018 and 2019 and reached 32 exabytes (equal to 3.2×10^{18} bytes) per month [23], which is primarily fueled by viewing multimedia content at increasingly higher resolution; see the mobile data traffic growth in Fig. 1.1. While there is increasing demand for wireless connectivity, the spectrum particularly between 500 MHz and 3 GHz is limited; and most of defined spectrum bands have already been allocated for governmental and commercial activities. Additionally, growing interest in the Internet of things (IoT) puts a strain on the available unlicensed spectral resources. Provided that the projected number of IoT devices in factories, businesses, and healthcare reaches 200 billion by 2020 [41], the currently available spectrum will be eventually overloaded and considerable interference issues will arise as a result.

Evident scarcity of spectral resources for (un)licensed bands has popularized mainly two different ideas as potential remedies: i) spectrum sharing; and ii) utilization of higher frequencies. As a manifestation of the former, cognitive radio networks (CRNs) have arguably gained center-stage prominence. Cognitive radios (CRs) are a set of devices equipped with cognition capabilities to learn the spatio-temporal and spectral

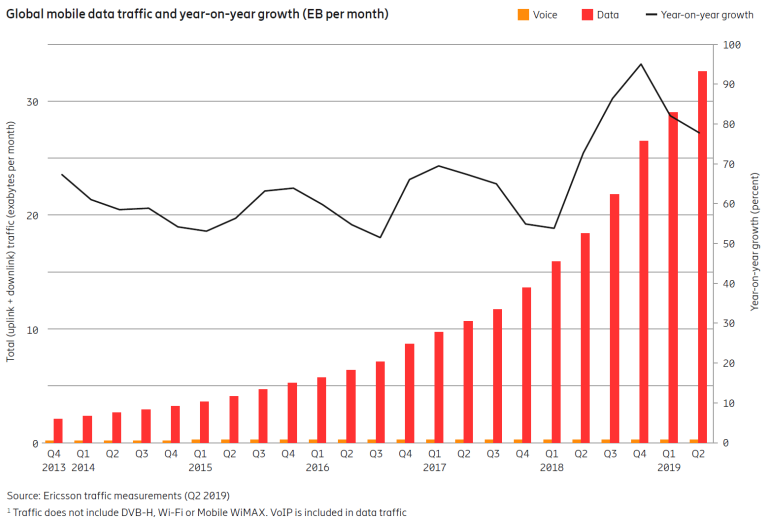


Figure 1.1: Global mobile data traffic and year-on-year growth [23].

usage patterns of near by users and networks. Such a level of cognition allows opportunistic utilization of the unused (un)licensed spectrum via spectrum sensing and dynamic spectrum access while avoiding interference in networks operating over the same band. This is particularly appealing in a recent situation that the licensed RF spectrum is often severely under-utilized depending on the time and location of communication [25]. While spectrum sharing has been proposed for more efficient use of existing spectral resources, utilization of higher frequencies addresses not only the spectral scarcity, but also increasing demand for higher date rates. Millimeter wave (mmWave) communications over the *licensed* spectrum between 30–300 GHz have recently gained more attention from the standards organization, the Federal Communications Commission (FCC), and academia as a means to bring "5G" cellular systems into the future, while those over the unlicensed spectrum were mainly studied to develop technologies for a personal area network (PAN) to deliver uncompressed high definition (HD) video, or standardized for a wireless local area network (WLAN); see e.g., WirelessHD [92] and IEEE 802.11ad [1], respectively.

While the aforementioned solutions promise more efficient utilization of the spectral resources and faster means of wireless communication, next-generation communication systems face formidable challenges as outlined next.

C1. Networks are ultra dense and heterogeneous. One differentiator of 5G networks relative to legacy generations (1–4G) is *heterogeneity*, which is induced by the convergence of "component" networks in various sizes operating over potentially different frequency bands. These so-termed heterogeneous networks (HetNets) are key enablers of 5G systems together with densification of the infrastructure to meet quality-of-service (QoS) expected by users and satisfy different service coverage requirements, provided that seamless interconnections among component networks are guaranteed. While CRs are considered as a key technology to accommodate the HetNets by providing adaptive handover between component networks [43], the successful operation of CRNs hinges critically on channel state information (CSI) over space, time, and frequency to find spectrum holes [47]. However, conventional point-to-point estimation methods such as ray-tracing [87, 93] do not provide feasible solutions for extremely dense and heterogeneous networks.

C2. Differences of mmWave channel relative to sub-6 GHz channel. Compared to the channel at sub-6 GHz, the millimeter wave channel shows significantly different characteristics due to the very short wavelength relative to the size of objects located in the propagation environment. This results in high sensitivity of signals to blockages, with consequently pronounced shadowing effects but relatively low diffraction [57]. For example, signal strength can be attenuated as much as 35dB by the human body [54]. Furthermore, the signal propagating over mmWave bands experiences higher path-loss than that at sub-6 GHz with omnidirectional antennas since the path-loss is inversely proportional to the wavelength squared by the Friis' law. In other words, mmWave communications become feasible through either co-siting with existing technologies, or directional transmissions and MIMO techniques with adaptive beamforming, to compensate for severe signal attenuation.

C3. New hardware constraints on massive MIMO for mmWave communication. To implement mmWave communication systems by addressing C2, it is inevitable to adopt MIMO techniques with antenna arrays having between 16 to 256 elements [37], which could be even larger at base stations in cellular networks. For such

a large number of antenna elements, several hardware constraints arise from a practical point of view. Conventional digital MIMO architectures at sub-6 GHz frequencies (generally with two antenna elements) entail a power amplifier (PA) and an RF chain with analog-to-digital converter (ADC), or digital-to-analog converter (DAC), associated with each antenna on top of all baseband connections. As the number of antenna elements increases, it becomes impractical to pack all these devices on a circuit board with limited space while placing antennas very close to each other to avoid granting lobes. Furthermore, power consumption is another critically limiting factor; e.g, power hungry devices such as ADC or PA consume 15–795 (mW) per antenna [27, 20]. Therefore, implementation of mmWave communications requires z beamforming architecture with low-power consumption while providing a sufficient spatial multiplexing gain by supporting a *massive* number of antenna arrays.

In this context, the present dissertation will leverage contemporary science and engineering tools from diverse disciplines in order to put forth analytical and algorithmic foundations to design and operate modern communication systems.

1.2 Channel-gain Cartography

The abiding goal of this thesis is to jointly address challenges C1–C3 under a principled machine learning framework. To tackle C1 and C2, we put forth algorithmic innovations for efficient and adaptive learning of global channel-state information for next-generation communication systems via *channel-gain cartography*. On the other hand, future research directions to address C3 will be discussed in Chapter 6.

Channel-gain cartography is a groundbreaking geostatistics-inspired application portraying the RF landscape impinging upon arbitrary spatial locations. The most appealing feature of this tool is the non-trivial capability of inferring channel-gain between arbitrary transceiver locations, even where no sensor is deployed, based only on measurements collected by a set of collaborating sensing radios. The vision of channel-gain cartography is to utilize the resulting channel-gain atlas for cross-layer design and assessment of the system-level performance of wireless networks; and to enhance hand-off, routing, interference management, and resource allocation, without requiring a large number of point-to-point channel estimates over wireless networks.

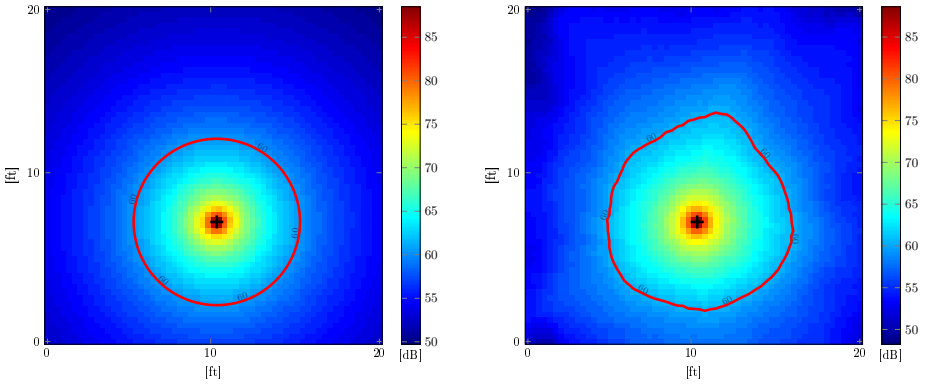
Channel-gain cartography leverages the notion of *spatial loss fields* (SLFs), which are maps quantifying the attenuation experienced by electromagnetic waves in radio frequency (RF) bands at every spatial position. The SLF model is used to estimate shadowing over an arbitrary radio link, and subsequently the associated channel gain as well. This enables construction of a map depicting a landscape of channel-gain from any point to a common end point in the region of interest. Considering that characterization of the propagation environment is critical for obtaining the channel-state information based on situational awareness, more accurate spectrum sensing and aggressive spatial reuse can be expected from utilization of a channel-gain map, instead of adopting a path-loss only model. Fig. 1.2 delineates spatial spectrum opportunity at a secondary receiver marked by a black cross, obtained via the proposed channel-gain map and the path-loss only model. For illustration purposes, the threshold to meet the QoS is set to 60dB and corresponding contour is drawn in red. Apparently, the spatial coverage of the receiver obtained by using the channel-gain map expands more than that by using the path-loss only model. This demonstrates nicely that more aggressive spatial reuse becomes available due to site-specific interference management through the proposed channel-gain map.

Such a non-trivial capability of inferring any-to-any channel-gain can be a key to success of spectrum reuse over HetNets, and co-siting for mmWave communications by enhancing hand-off, routing, and interference management. These considerations motivate the innovative machine learning and Bayesian inference algorithms for channel-gain cartography that will be developed in the following chapters and, accordingly, a significant departure from conventional per-link channel-gain and interference level estimation will be advocated.

1.3 Thesis Outline

The remainder of the thesis is organized as follows.

Chapter 2 reviews channel-gain cartography. The concept of channel-gain cartography is introduced with its functionality. Prior works including radio tomography are reviewed as well. Afterwards, the system model and problem statement are presented, which are considered throughout the thesis.



(a) (b)

Figure 1.2: Spatial spectrum opportunity of a CR, obtained via (a) a path-loss only model; and (b) a channel-gain map.

Chapter 3 puts forth channel-gain cartography leveraging low-rank and sparsity, having as goal to construct a channel-gain map with a relatively small number of measurements. The key idea is to postulate that the SLF has a low-rank structure potentially corrupted by sparse outliers. Such a model is particularly appealing for urban and indoor propagation scenarios, where regular placement of buildings and walls renders a scene inherently of low-rank, while sparse outliers can pick up the artifacts that do not conform to the low-rank model. We develop an efficient batch algorithm as well as its online version via stochastic approximation (SA) [84, 48]. Performance of the proposed algorithms is evaluated with a rigorous performance analysis and extensive numerical tests on synthetic and real datasets.

Chapter 4 introduces a novel Bayesian framework for channel-gain cartography. To take into account spatial heterogeneity of the propagation environment when learning the SLF, we propose a two-layer Bayesian SLF model based on a binary hidden Markov random field along with Markov chain Mote Carlo (MCMC) methods for inference [30]. Besides accounting for heterogeneous propagation environments, another contribution here is a data-adaptive sensor selection technique, with the goal of reducing SLF uncertainty, by cross-fertilizing ideas from the fields of experimental design [26] and active learning [55]. Efficacy of the proposed solution is established through extended synthetic and real data tests.

Chapter 5 builds on the algorithms and results of Chapter 4, and devises a variational Bayes approach to adaptive Bayesian channel-gain cartography. The aforementioned Bayesian SLF model is generalized first by adopting a K-ary hidden Markov random field, to address a richer class of environmental heterogeneity. Subsequently, variational Bayes (VB) algorithms are developed to provide efficient field estimators at affordable complexity. To bypass a novel but intractable sensor selection criterion, its efficient proxy can be obtained thanks to the availability of an approximate posterior model from the proposed VB algorithm. Numerical tests on synthetic and real data corroborate the effectiveness of the proposed algorithms.

Finally, Chapter 6 presents a concluding discussion of the proposed approaches, along with future research directions.

1.4 Notational Conventions

The following notation is used throughout the subsequent chapters. Bold uppercase (lowercase) letters denote matrices (column vectors). Calligraphic letters are used for sets; $\mathbf{I}_n$ is the $n \times n$ identity matrix; $\mathbf{0}_n$ denotes an $n \times 1$ vector of all zeros, and $\mathbf{0}_{n\times n}$ an $n \times n$ matrix of all zeros. Operators $(\cdot)^\top$, $\mathrm{tr}(\cdot)$, $\sigma_i(\cdot)$, and $\lambda_{\max}(\cdot)$ represent the transposition, trace, the i-th largest singular value, and the largest eigenvalue of a matrix, respectively; $|\cdot|$ is used for the cardinality of a set, the magnitude of a scalar, and the determinant of a matrix. $\mathbf{R} \succeq \mathbf{0}$ signifies that $\mathbf{R}$ is positive semidefinite. The ℓ_1-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is $\|\mathbf{X}\|_1 := \sum_{i,j=1}^n |X_{ij}|$. The ℓ_∞-norm of $\mathbf{X} \in \mathbb{R}^{n\times n}$ is represented by $\|\mathbf{X}\|_\infty := \max\{|X_{ij}| : i, j = 1, \ldots, n\}$. For two matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$, the matrix inner product is $\langle \mathbf{X}, \mathbf{Y} \rangle := \mathrm{tr}(\mathbf{X}^\top \mathbf{Y})$. The Frobenius norm of matrix $\mathbf{Y}$ is $\|\mathbf{Y}\|_F := \sqrt{\mathrm{tr}(\mathbf{Y}\mathbf{Y}^\top)}$. The spectral norm of $\mathbf{Y}$ is $\|\mathbf{Y}\| := \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Y}\mathbf{x}\|_2$, and $\|\mathbf{Y}\|_* := \sum_i \sigma_i(\mathbf{Y})$ is the nuclear norm of $\mathbf{Y}$. For a function $h : \mathbb{R}^{m\times n} \to \mathbb{R}$, the directional derivative of h at $\mathbf{X} \in \mathbb{R}^{m\times n}$ along a direction $\mathbf{D} \in \mathbb{R}^{m\times n}$ is denoted by $h'(\mathbf{X}; \mathbf{D}) := \lim_{t\to 0+}[h(\mathbf{X} + t\mathbf{D}) - h(\mathbf{X})]/t$. $\mathrm{vec}(\mathbf{X})$ produces a column vector $\mathbf{x} \in \mathbb{R}^{mn}$ by stacking the columns of a matrix one after the other ($\mathrm{unvec}(\mathbf{x})$ denotes the reverse process). For a vector $\mathbf{y} \in \mathbb{R}^n$ and an $n \times n$ weight matrix $\mathbf{\Delta}$, the weighted norm of $\mathbf{y}$ is $\|\mathbf{y}\|^2_{\mathbf{\Delta}} := \mathbf{y}^\top \mathbf{\Delta} \mathbf{y}$.

Chapter 2

Channel-gain Cartography

2.1 Preliminaries and Motivation

Conventional acquisition of the channel-state information (CSI) on a per-link basis might become inadequate for emerging wireless technologies, since needs for accounting situational awareness are unrelentingly demanded to accomplish dynamic spectral resource control for spectrum sharing in next-generation communication systems [96, 75, 40].

To meet the demands for tools enabling aggressive and full opportunistic utilization of the unused (un)licensed spectrum, radio frequency (RF) cartography was proposed as an instrumental concept originally for cognitive radios (CRs) [46]. Based on the measurements collected by spatially distributed sensing radios, RF cartography provides tools to construct maps over the space, time, and frequency, portraying a RF landscape in which a CR network is deployed. Notable RF maps that have been proposed include a *power spectral density (PSD) map*, which acquires the ambient interference power distribution, revealing the crowded regions that CR transceivers need to avoid [6]; and a *channel-gain (CG) map*, which delineates the spatial distribution of channel-gain in a given geographical region through a collaborative network of CRs, allowing CR networks to perform accurate spectrum sensing and aggressive spatial reuse [47]. The present thesis focuses on channel-gain cartography.

Given channel-gain measurements from sensing radios at known locations in a region of interest, the goal of channel-gain cartography is to estimate or predict channel-gain from *any* point to a deployed radio, which is henceforth termed as a local CG map; as

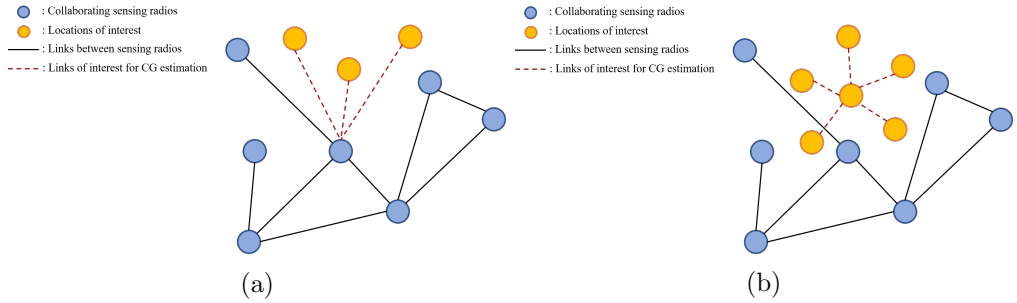


Figure 2.1: Illustration of channel-gain maps: (a) local; and (b) global maps.

well as that of an arbitrary wireless link from *any* point to *any* other point in space, (i.e., links not having communication ending points in common with the links between existing sensing radio pairs), which constitute a so-termed global CG map: local and global CG maps are illustrated in Fig. 2.1. The vision of channel-gain cartography is to utilize resultant channel-gain atlas for cross-layer design and assessment of the system-level performance of wireless networks; and to provide the vital information for interference management, resource allocation, and spectrum sensing. It is also notable that channel-gain cartography relies on *incoherent* measurements containing no phase information, e.g., the received signal strength (RSS). Such simplification saves costs for synchronization needed to calibrate phase differences among waveforms received at different sensors.

The key premise behind channel-gain cartography is that spatially close radio links exhibit similar shadowing due to the presence of common obstructions. This shadowing correlation is related to the geometry of objects present in the area that waves propagate through [71, 2]. As a result, shadowing is modeled as the weighted line integral of the underlying two-dimensional *spatial loss field* (SLF), which is a map quantifying the attenuation experienced by electromagnetic waves in RF bands at every spatial position [71]. The weights in the integral are determined by a function depending on the transmitter-receiver locations [71, 33, 82], which models the SLF effect on shadowing over a link. Inspired by this SLF model, linear interpolation techniques such as kriging were further employed to estimate shadowing based on spatially correlated measurements [18], while spatio-temporal dynamics were tracked via Kalman filtering [47]. Instead of relying on heuristic criteria to choose the weight function, [82] provides blind

algorithms to learn the weight function using a non-parametric kernel regression, while estimating the SLF via regularized least-squares (LS) methods. Note that another body of work leveraging the SLF model is radio tomographic imaging (RTI) [91]. Benefiting from the ability of RF waves to penetrate physical structures such as trees and buildings, RTI provides a means of device-free passive localization [94, 97], and has found diverse applications in disaster response for e.g., detecting individuals trapped in buildings or smoke [90]. To detect locations of changes in the propagation environment, one can use the difference between the SLF across consecutive time slots [91, 89]. To cope with multipath in a cluttered environment, multi-channel measurements can be utilized to enhance localization accuracy [44]. Although these are calibration-free approaches, they cannot reveal static objects in the area of interest. It is also possible to replace the SLF with a label field indicating presence (or absence) of objects in motion on each voxel [90], and leverage the influence that moving objects on the propagation path have, on the variance of a RSS measurement. On the other hand, the SLF itself was reconstructed in [32, 33] to depict static objects in the area of interest, but calibration was necessary by using extra measurements (e.g., collected in free space). One can avoid extra data for calibration by estimating the SLF together with pathloss components [8, 82]. Exploiting the sparse occupancy of the target objects in a monitored area, sparsity-leveraging algorithms for constructing obstacle maps were also developed [66, 45, 65].

The overarching contribution of the present thesis is to develop algorithmic foundations for effective data-driven channel learning by capitalizing on the inherent structure of measurement data, rather than relying heavily on the physics of RF propagation. RF propagation environment is particularly taken into consideration as the prior information to learn the shadowing model, inspired by a fact that absorption captured by the SLF allows one to discern objects located in the area of interest. We propose two SLF models: i) a low-rank plus sparse matrix model [16, 24, 59]; and ii) a hidden Markov random field (MRF) model [38]. The former is appealing for urban and indoor propagation scenarios, where regular placement of buildings and walls renders a scene inherently of low rank, while sparse outliers can pick up the artifacts that do not conform to the low-rank model. On the other hand, the latter is useful when the propagation environment is spatially diverse due to a combination of free space and objects in different sizes and materials, which subsequently induces statistical heterogeneity in the SLF. Efficient

solution methods leveraging aforementioned SLF models will be developed, and their efficacy is shown through extensive synthetic and real data tests.

2.2 System Model and Problem Statement

Consider a set of sensors deployed over a two-dimensional geographical area $\mathcal{A} \subset \mathbb{R}^2$. After averaging out small-scale fading effects, the channel-gain measurement over a link between a transmitter located at $\mathbf{x} \in \mathcal{A}$ and a receiver located at $\mathbf{x}' \in \mathcal{A}$ can be represented (in dB) as

$$g(\mathbf{x}, \mathbf{x}') = g_0 - \gamma 10 \log_{10} d(\mathbf{x}, \mathbf{x}') - s(\mathbf{x}, \mathbf{x}') \tag{2.1}$$

where g_0 is the path gain at unit distance; $d(\mathbf{x}, \mathbf{x}') := \|\mathbf{x} - \mathbf{x}'\|$ is the Euclidean distance between the transceivers at $\mathbf{x}$ and $\mathbf{x}'$; γ is the pathloss exponent; and $s(\mathbf{x}, \mathbf{x}')$ is the attenuation due to shadow fading.

A tomographic shadow fading model is [71, 33, 51]

$$s(\mathbf{x}, \mathbf{x}') = \int_{\mathcal{A}} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \tag{2.2}$$

where $f : \mathcal{A} \to \mathbb{R}$ denotes the *spatial loss field* (SLF) capturing the attenuation at location $\tilde{\mathbf{x}}$, and $w : \mathcal{A} \times \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ is a weight function describing how the SLF at $\tilde{\mathbf{x}}$ contributes to the shadowing experienced over the link $\mathbf{x}$–$\mathbf{x}'$. Typically, w confers a greater weight $w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}})$ to those locations $\tilde{\mathbf{x}}$ lying closer to the link $\mathbf{x}$–$\mathbf{x}'$. Examples of the weight function include the *normalized ellipse model* [89]

$$w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) := \begin{cases} 1/\sqrt{d(\mathbf{x}, \mathbf{x}')}, & \text{if } d(\mathbf{x}, \tilde{\mathbf{x}}) + d(\mathbf{x}', \tilde{\mathbf{x}}) < d(\mathbf{x}, \mathbf{x}') + \lambda/2 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

where $\lambda > 0$ is a tunable parameter. The value of λ is commonly set to the wavelength to assign non-zero weights only within the first Fresnel zone. It is worth to mention that the weight function can be learned via a non-parametric kernel regression, instead of relying on on heuristic criteria to choose the weight function; see [82] for details. In

practice, the integral in (2.2) is approximated by a finite sum as

$$s(\mathbf{x},\mathbf{x}') \simeq c\sum_{i=1}^{N_g} w(\mathbf{x},\mathbf{x}',\tilde{\mathbf{x}}_i) f(\tilde{\mathbf{x}}_i) \tag{2.4}$$

where $\{\tilde{\mathbf{x}}_i\}_{i=1}^{N_g}$ is a grid of points over $\mathcal{A}$ and c is a constant that can be set to unity without loss of generality by absorbing any scaling factor in f. Clearly, (2.4) shows that $s(\mathbf{x},\mathbf{x}')$ depends on f only through its values at the grid points.

The model in (2.2) describes how the spatial distribution of obstructions in the propagation path influences the attenuation between a pair of locations. The usefulness of (2.2) is twofold: i) as f represents absorption across space, it can be used for imaging; and ii) once f and w are known, the gain between any two points $\mathbf{x}$ and $\mathbf{x}'$ can be recovered through (2.1) and (2.2), which is precisely the objective of channel-gain cartography.

All in all, the objective of channel-gain cartography is tantamount to estimating f. To this end, N sensors located at $\{\mathbf{x}_1,\ldots,\mathbf{x}_N\} \in \mathcal{A}$ collaboratively obtain channel-gain measurements. At time slot τ, the radios indexed by $n(\tau)$ and $n'(\tau)$ measure the channel-gain $\breve{g}_\tau := g(\mathbf{x}_{n(\tau)},\mathbf{x}_{n'(\tau)}) + \nu_\tau$ by exchanging training sequences known to both transmitting and receiving radios, where $n(\tau), n'(\tau) \in \{1,\ldots,N\}$ and ν_τ denotes measurement noise. It is supposed that g_0 and γ have been estimated during a calibration stage. After subtracting known components from $\breve{g}_\tau$, the shadowing estimate is found as

$$\begin{aligned}\breve{s}_\tau &:= g_0 - \gamma 10\log_{10} d(\mathbf{x}_{n(\tau)},\mathbf{x}_{n'(\tau)}) - \breve{g}_\tau \\ &= s(\mathbf{x}_{n(\tau)},\mathbf{x}_{n'(\tau)}) - \nu_\tau. \end{aligned}\tag{2.5}$$

Having available $\breve{\mathbf{s}}_t := [\breve{s}_1,\ldots,\breve{s}_t]^\top \in \mathbb{R}^t$ along with the known set of links $\{(\mathbf{x}_{n(\tau)},\mathbf{x}_{n'(\tau)})\}_{\tau=1}^t$ and the weight function w at the fusion center, the problem is to estimate f, and thus $\boldsymbol{f} := [f(\tilde{\mathbf{x}}_1),\ldots,f(\tilde{\mathbf{x}}_{N_g})]^\top \in \mathbb{R}^{N_g}$ using (2.4). Once $\hat{\boldsymbol{f}}$ is obtained, shadowing and subsequently channel-gain across any link $\mathbf{x}$–$\mathbf{x}'$ can be estimated via (2.4) and (2.1) as

$$\hat{s}(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{N_g} w(\mathbf{x},\mathbf{x}',\tilde{\mathbf{x}}_i)\hat{f}(\tilde{\mathbf{x}}_i) \tag{2.6}$$

$$\hat{g}(\mathbf{x}, \mathbf{x}') = g_0 - \gamma 10 \log_{10} d(\mathbf{x}, \mathbf{x}') - \hat{s}(\mathbf{x}, \mathbf{x}'). \tag{2.7}$$

Chapter 3

Channel-gain Cartography leveraging Low-rank and Sparsity

A task of channel-gain cartography is well-motivated to benefit operation of cognitive radio networks (CRNs) by providing a means of site-specific interference management and subsequently, spectrum sensing. Although more sophisticated methodologies for channel modeling do exist including ray tracing [87, 93] to serve the same purpose, the computational cost and requirements on various structural/geometric prior information may hinder their use in CR applications. Capitalizing on experimentally validated notion of the spatial loss field (SLF) [2], we will provide a computationally efficient solution by leveraging the inherent structure of data, rather than relying heavily on the physics of radio frequency (RF) signal propagation.

Our work interpolates the channel gains based on the SLF reconstructed from a small number of measurements using a low-rank and sparse matrix model. The key idea is to postulate that the SLF has a low-rank structure potentially corrupted by sparse outliers. Such a model is particularly appealing for urban and indoor propagation scenarios, where regular placement of buildings and walls renders a scene inherently of low rank, while sparse outliers can pick up the artifacts that do not conform to the low-rank model. While it is true that urban and indoor environments have distinct profiles due to the different scales and density of obstacles, our data model can capture the *structural regularity* of obstacles, possibly at different scales, as validated through

synthetic and real data examples in Section 3.3. The sparse term helps robustify this model by filtering out the measurements that do not conform to the low-rank structure. This is essentially the idea behind robust principal component analysis [16], which is a powerful data model that has been used widely.

In fact, since a shadowing measurement is modeled as a linear tomographic measurement of the SLF, the map recovery task reduces to an instance of compressive principal component pursuit (CPCP) [95]. In general, the CPCP problem recovers the low-rank and sparse matrices from a small set of linearly projected measurements. Our algorithms are applicable to this general problem class.

We develop efficient batch and online algorithms for channel-gain cartography. By replacing the nuclear norm-based regularizer with a bi-factorization surrogate, a block coordinate descent (BCD) algorithm becomes available to avoid costly singular value decomposition (SVD) per iteration. Although the resulting optimization problem is non-convex, the batch solver can attain the global optimum under appropriate conditions. For the online algorithm, a stochastic successive upper-bound minimization strategy is adopted, leading to a stochastic gradient descent (SGD) update rule, which enjoys low computational complexity. The iterates generated by the online algorithm are provably convergent to the stationary point of the batch problem.

3.1 Channel-gain estimation using Low-rank and Sparsity

The goal of the present section is to estimate the SLF by leveraging its inherent low-rank and sparse attribute. To this end, let matrix $\mathbf{F} := \text{unvec}(\boldsymbol{f}) \in \mathbb{R}^{N_x \times N_y}$ denote the SLF, sampled by the N_x-by-N_y grid, where $N_g = N_x N_y$. Let further define $\mathbf{w}_{nn'}^{(t)} := [w(\mathbf{x}_{n(t)}, \mathbf{x}_{n'(t)}, \tilde{\mathbf{x}}_1), \ldots, w(\mathbf{x}_{n(t)}, \mathbf{x}_{n'(t)}, \tilde{\mathbf{x}}_{N_g})]^\top \in \mathbb{R}^{N_g}$. Then, the weight matrix $\mathbf{W}_{nn'}^{(t)} := \text{unvec}(\mathbf{w}_{nn'}^{(t)}) \in \mathbb{R}^{N_x \times N_y}$ corresponding to link $\mathbf{x}_{n(t)}$–$\mathbf{x}_{n'(t)}$ is constructed in similar manner. Subsequently, the shadow fading over link $\mathbf{x}_{n(t)}$–$\mathbf{x}_{n'(t)}$ in (2.4) can be expressed as a linear projection of the SLF given by

$$s(\mathbf{x}_{n(t)}, \mathbf{x}_{n'(t)}) \simeq \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{F} \rangle. \tag{3.1}$$

In the following sections of Chapter 3, the measurement model in (3.1) will be specifically considered.

3.1.1 CPCP Problem formulation

The low-rank plus sparse structure has been advocated in various problems in machine learning and signal processing [16, 24, 59]. Low-rank matrices are effective in capturing slow variation or regular patterns, and sparsity is instrumental for incorporating robustness against outliers. Inspired by these, we postulate that $\mathbf{F}$ has a low-rank-plus-sparse structure as

$$\mathbf{F} = \mathbf{L} + \mathbf{E} \tag{3.2}$$

where matrix $\mathbf{L}$ is low-rank, and $\mathbf{E}$ is sparse. This model is particularly attractive in urban or indoor scenarios where the obstacles often possess regular patterns, while the sparse term can capture irregularities that do not conform to the low-rank model.

Redefine $\check{s}_{nn'}^{(t)} := \check{s}(\mathbf{x}_{n(t)}, \mathbf{x}_{n'(t)})$ for brevity. Let $\mathcal{M}^{(t)}$ be the set of links, for which channel gain measurements are made at time t, and collect those measurements in vector $\check{\mathbf{s}}^{(t)} \in \mathbb{R}^{|\mathcal{M}^{(t)}|}$. Toward estimating $\mathbf{F}^{(t)}$ that obeys (3.2), consider the cost

$$c^{(t)}(\mathbf{L}, \mathbf{E}) := \frac{1}{2} \sum_{(n,n') \in \mathcal{M}^{(t)}} \left(\langle \mathbf{W}_{nn'}^{(t)}, \mathbf{L} + \mathbf{E} \rangle - \check{s}_{nn'}^{(t)} \right)^2 \tag{3.3}$$

which fits the shadowing measurements to the model. Then, with T denoting the total number of time slots taking measurements, we adopt the following optimization criterion

$$\text{(P1)} \quad \min_{\mathbf{L}, \mathbf{E} \in \mathbb{R}^{N_x \times N_y}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{L}, \mathbf{E}) + \mu_{\mathbf{L}} \|\mathbf{L}\|_* + \mu_{\mathbf{E}} \|\mathbf{E}\|_1 \right] \tag{3.4}$$

where $\beta \in (0, 1]$ is the forgetting factor that can be optionally put in to weigh the recent observations more heavily. The nuclear norm regularization term promotes a low-rank $\mathbf{L}$, while the ℓ_1-norm encourages sparsity in $\mathbf{E}$. Parameters $\mu_{\mathbf{L}}$ and $\mu_{\mathbf{E}}$ are appropriately chosen to control the effect of these regularizers. Conditions for exact recovery through a related convex formulation in the absence of measurement noise can be found in [95].

Problem (3.4) is convex, and can be tackled using existing efficient solvers, such as the interior-point method. Once the optimal $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{E}}$ are found, the desired $\widehat{\mathbf{F}}$ is obtained as $\widehat{\mathbf{F}} = \widehat{\mathbf{L}} + \widehat{\mathbf{E}}$. However, the general-purpose optimization packages tend to scale poorly as the problem size grows. Specialized algorithms developed for related problems

often employ costly SVD operations iteratively [95]. Furthermore, such an algorithm might not be amenable for an online implementation. Building on [58] and [81], an efficient solution is proposed next with reduced complexity.

3.1.2 Efficient batch solution

Without loss of generality, consider replacing $\mathbf{L}$ with the low-rank product $\mathbf{P}\mathbf{Q}^\top$, where $\mathbf{P} \in \mathbb{R}^{N_x \times \rho}$ and $\mathbf{Q} \in \mathbb{R}^{N_y \times \rho}$, and ρ is a pre-specified overestimate of the rank of $\mathbf{L}$. It is known that (e.g., [81])

$$\begin{aligned} \|\mathbf{L}\|_* = \min_{\mathbf{P},\mathbf{Q}} \frac{1}{2}\left(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2\right) \\ \text{subject to } \mathbf{L} = \mathbf{P}\mathbf{Q}^\top. \end{aligned} \tag{3.5}$$

Thus, a natural re-formulation of (3.4) is (see also [58])

$$\text{(P2)} \quad \min_{\mathbf{P},\mathbf{Q},\mathbf{E}} f(\mathbf{P},\mathbf{Q},\mathbf{E}) := \sum_{\tau=1}^{T} \beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^\top,\mathbf{E}) + \frac{\mu_{\mathbf{L}}}{2}\left(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2\right) + \mu_{\mathbf{E}}\|\mathbf{E}\|_1\right]. \tag{3.6}$$

Instead of seeking the $N_x N_y$ entries of $\mathbf{L}$, the factorization approach (3.6) entails only $(N_x + N_y)\rho$ unknowns, thus reducing complexity and memory requirements significantly when $\rho \ll \min\{N_x, N_y\}$. Furthermore, adoption of the separable Frobenius norm regularizer in (P2) comes with no loss of optimality as asserted in the following lemma.

Lemma 1: *If $\{\widehat{\mathbf{L}}, \widehat{\mathbf{E}}\}$ minimize (P1) and we choose $\rho \geq$ rank$(\widehat{\mathbf{L}})$, then, (P2) is equivalent to (P1) at the minimum.*

Proof: It is clear that the minimum of (P1) is no larger than that of

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{E}} \sum_{\tau=1}^{T} \beta^{T-\tau}\left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^\top,\mathbf{E}) + \mu_{\mathbf{L}}\|\mathbf{P}\mathbf{Q}^\top\|_* + \mu_{\mathbf{E}}\|\mathbf{E}\|_1\right] \tag{3.7}$$

since the search space is reduced by the reparameterization $\mathbf{L} = \mathbf{P}\mathbf{Q}^\top$ with $\rho \leq \min\{N_x, N_y\}$. Now (3.5) implies that the minimum of (3.7) is no larger than that of (P2). However, the inequality is tight since the objectives of (P1) and (P2) are identical for $\mathbf{E} := \widehat{\mathbf{E}}$, $\mathbf{P} := \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}^{1/2}$, and $\mathbf{Q} := \widehat{\mathbf{V}}\widehat{\mathbf{\Sigma}}^{1/2}$, where $\widehat{\mathbf{L}} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^\top$ is the SVD.

Consequently, (P1) and (P2) have identical costs at the minimum. ■

Although (P1) is a convex optimization problem, (P2) is not. Thus, in general, one can obtain only a *locally* optimal solution of (P2), which may not be the globally optimal solution of (P1). Interestingly, under appropriate conditions, global optimality can be guaranteed for the local optima of (P2), as claimed in the following proposition.

Proposition 1: *If* $\{\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}\}$ *is a stationary point of (P2),* $\bar{\beta} := \sum_{\tau=1}^{T} \beta^{T-\tau}$*, and* $\|\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}})\| \leq \mu_{\mathbf{L}}\bar{\beta}$ *with*

$$\tilde{f}(\widehat{\mathbf{L}}, \widehat{\mathbf{E}}) := \sum_{\tau=1}^{T} \beta^{T-\tau} \left[\sum_{(n,n')\in\mathcal{M}^{(\tau)}} \left(\langle \mathbf{W}_{nn'}^{(\tau)}, \widehat{\mathbf{L}} + \widehat{\mathbf{E}} \rangle - \check{s}_{nn'}^{(\tau)} \right) \mathbf{W}_{nn'}^{(\tau)} \right] \tag{3.8}$$

then $\{\widehat{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \widehat{\mathbf{E}} := \bar{\mathbf{E}}\}$ *is a globally optimal solution to (P1).*
Proof: See Appendix A.1.

A stationary point of (P2) can be obtained through a block coordinate-descent (BCD) algorithm, where the optimization is performed in a cyclic fashion over one of $\{\mathbf{E}, \mathbf{P}, \mathbf{Q}\}$ with the remaining two variables fixed. In fact, since the term $\mu_{\mathbf{E}}\|\mathbf{E}\|_1$ is separable in the individual entries as well, the cyclic update can be stretched all the way up to the individual entries of $\mathbf{E}$ without affecting convergence [86]. The proposed solver entails an iterative procedure comprising three steps per iteration $k = 1, 2, \ldots$

[S1] Update E:

$$\mathbf{E}[k+1] = \arg\min_{\mathbf{E}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}[k]\mathbf{Q}^\top[k], \mathbf{E}) + \mu_{\mathbf{E}}\|\mathbf{E}\|_1 \right]$$

[S2] Update P:

$$\mathbf{P}[k+1] = \arg\min_{\mathbf{P}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}\mathbf{Q}^\top[k], \mathbf{E}[k+1]) + \frac{\mu_{\mathbf{L}}}{2}\|\mathbf{P}\|_F^2 \right]$$

[S3] Update Q:

$$\mathbf{Q}[k+1] = \arg\min_{\mathbf{Q}} \sum_{\tau=1}^{T} \beta^{T-\tau} \left[c^{(\tau)}(\mathbf{P}[k+1]\mathbf{Q}^\top, \mathbf{E}[k+1]) + \frac{\mu_{\mathbf{L}}}{2}\|\mathbf{Q}\|_F^2 \right].$$

To update each block variable, the cost in (P2) is minimized while fixing the other block variables to their up-to-date iterates. To detail the update rules, let $\boldsymbol{\mathcal{W}}^{(t)} \in \mathbb{R}^{N_x N_y \times |\mathcal{M}(t)|}$ be a matrix with columns equal to $\mathbf{w}_{nn'}^{(t)}$ for $(n, n') \in \mathcal{M}(t)$. Define $\boldsymbol{\mathcal{W}} := [\sqrt{\beta^{T-1}}\boldsymbol{\mathcal{W}}^{(1)} \ldots \sqrt{\beta^0}\boldsymbol{\mathcal{W}}^{(T)}]$, $\check{\mathbf{s}} := [\sqrt{\beta^{T-1}}\check{\mathbf{s}}^{(1)\top} \ldots \sqrt{\beta^0}\check{\mathbf{s}}^{(T)\top}]^\top$, and $\mathbf{e} := \mathrm{vec}(\mathbf{E})$. Then,

one can write $\sum_{\tau=1}^{T}\beta^{T-\tau}c^{(\tau)}(\mathbf{P}\mathbf{Q}^\top,\mathbf{E})=\|\boldsymbol{\mathcal{W}}^\top\mathrm{vec}(\mathbf{P}\mathbf{Q}^\top+\mathbf{E})-\check{\mathbf{s}}\|_2^2$. Let e_l denote the l-th entry of $\mathbf{e}$, and $\mathbf{e}_{-l}$ represent the replica of $\mathbf{e}$ without its l-th entry. Similarly, let $\boldsymbol{\omega}_l^\top$ denote the l-th row of the matrix $\boldsymbol{\mathcal{W}}$, and $\boldsymbol{\mathcal{W}}_{-l}$ denote the matrix $\boldsymbol{\mathcal{W}}$ with its l-th row removed. The soft-thresholding function $\mathrm{soft\_th}(\cdot;\mu_{\mathbf{E}})$ is defined as

$$\mathrm{soft\_th}(x;\mu_{\mathbf{E}}):=\mathrm{sign}(x)\max\{0,|x|-\mu_{\mathbf{E}}\}. \tag{3.9}$$

Minimization in **[S1]** proceeds sequentially over the individual entries of $\mathbf{e}$. At iteration k, each entry is updated via

$$e_l[k+1]=\arg\min_{e_l}\frac{1}{2}\|e_l\boldsymbol{\omega}_l-\check{\mathbf{s}}\|_2^2+\mu_{\mathbf{E}}\bar{\beta}|e_l|,\ l=1,\ldots,N_xN_y \tag{3.10}$$

where $\check{\mathbf{s}}_l[k]:=\check{\mathbf{s}}-\boldsymbol{\mathcal{W}}^\top\mathrm{vec}(\mathbf{P}[k]\mathbf{Q}^\top[k])-\boldsymbol{\mathcal{W}}_{-l}^\top\mathbf{e}_{-l}$. The closed-form solution for e_l is obtained as

$$e_l[k+1]=\frac{\mathrm{soft\_th}(\boldsymbol{\omega}_l^\top\check{\mathbf{s}}_l[k];\mu_{\mathbf{E}}\bar{\beta})}{\|\boldsymbol{\omega}_l\|_2^2}. \tag{3.11}$$

Matrices $\mathbf{P}$ and $\mathbf{Q}$ are similarly updated over their rows through **[S2]** and **[S3]**. Let $\mathbf{p}_i$ be the i-th row of $\mathbf{P}$, transposed to a column vector; i.e., $\mathbf{P}:=[\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_{N_x}]^\top$. Define $\tilde{\boldsymbol{\mathcal{W}}}_i^{(t)}\in\mathbb{R}^{|\mathcal{M}(t)|\times N_y}$ to be the matrix whose rows are the i-th rows of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n')\in\mathcal{M}(t)}$ denoted as $\tilde{\mathbf{w}}_{nn',i}^{(t)\top}$, and $\tilde{\mathbf{s}}_i^{(t)}\in\mathbb{R}^{|\mathcal{M}(t)|}$ a vector with entries equal to

$$\tilde{s}_{nn',i}^{(t)}:=\check{s}_{nn'}^{(t)}-\langle\mathbf{W}_{nn'}^{(t)},\mathbf{E}[k+1]\rangle-\sum_{j\neq i}^{N_x}\tilde{\mathbf{w}}_{nn',j}^{(t)\top}\mathbf{Q}[k]\mathbf{p}_j \tag{3.12}$$

for $(n,n')\in\mathcal{M}(t)$. Define also $\tilde{\boldsymbol{\mathcal{W}}}_i:=[\sqrt{\beta^{T-1}}\tilde{\boldsymbol{\mathcal{W}}}_i^{(1)\top}\ldots\sqrt{\beta^0}\tilde{\boldsymbol{\mathcal{W}}}_i^{(T)\top}]^\top$ and $\tilde{\mathbf{s}}_i:=[\sqrt{\beta^{T-1}}\tilde{\mathbf{s}}_i^{(1)\top}$
$\ldots\sqrt{\beta^0}\tilde{\mathbf{s}}_i^{(T)\top}]^\top$. Then, $\mathbf{p}_i$ is updated by solving a ridge-regression problem as

$$\mathbf{p}_i[k+1]=\arg\min_{\mathbf{p}_i}\left[\frac{1}{2}\|\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k]\mathbf{p}_i-\tilde{\mathbf{s}}_i\|_2^2+\frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\|\mathbf{p}_i\|_2^2\right]$$

whose solution is given in closed form by

$$\mathbf{p}_i[k+1] = \left[\mathbf{Q}^\top[k]\tilde{\boldsymbol{\mathcal{W}}}_i^\top\tilde{\boldsymbol{\mathcal{W}}}_i\mathbf{Q}[k] + \mu_{\mathbf{L}}\bar{\beta}\mathbf{I}_\rho\right]^{-1}\mathbf{Q}^\top[k]\tilde{\boldsymbol{\mathcal{W}}}_i^\top\tilde{\mathbf{s}}_i \tag{3.13}$$

which involves matrix inversion of dimension only ρ-by-ρ. Likewise, let $\mathbf{q}_i$ denote the i-th row of $\mathbf{Q}$, transposed to a column vector; i.e., $\mathbf{Q} := [\mathbf{q}_1, \ldots, \mathbf{q}_{N_y}]^\top$. Define also $\breve{\boldsymbol{\mathcal{W}}}_i := [\sqrt{\beta^{T-1}}\breve{\boldsymbol{\mathcal{W}}}_i^{(1)\top} \ldots \sqrt{\beta^0}\breve{\boldsymbol{\mathcal{W}}}_i^{(T)\top}]^\top$ and $\breve{\mathbf{s}}_i := [\sqrt{\beta^{T-1}}\breve{\mathbf{s}}_i^{(1)\top} \ldots \sqrt{\beta^0}\breve{\mathbf{s}}_i^{(T)\top}]^\top$, where $\breve{\boldsymbol{\mathcal{W}}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|\times N_x}$ is the matrix whose rows are the transpositions of the i-th columns of $\{\mathbf{W}_{nn'}^{(t)}\}_{(n,n')\in\mathcal{M}(t)}$, denoted as $\breve{\mathbf{w}}_{nn',i}^{(t)}$, and $\breve{\mathbf{s}}_i^{(t)} \in \mathbb{R}^{|\mathcal{M}(t)|}$ has entries

$$\breve{s}_{nn',i}^{(t)} := \breve{s}_{nn'}^{(t)} - \langle \mathbf{W}_{nn'}^{(t)}, \mathbf{E}[k+1]\rangle - \sum_{j\neq i}^{N_y} \breve{\mathbf{w}}_{nn',j}^{(t)\top}\mathbf{P}[k+1]\mathbf{q}_j \tag{3.14}$$

for $(n, n') \in \mathcal{M}(t)$. The update for $\mathbf{q}_i$ is then given by solving another ridge regression problem to obtain

$$\mathbf{q}_i[k+1] = \arg\min_{\mathbf{q}_i}\left[\frac{1}{2}\|\breve{\boldsymbol{\mathcal{W}}}_i\mathbf{P}[k+1]\mathbf{q}_i - \breve{\mathbf{s}}_i\|_2^2 + \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\|\mathbf{q}_i\|_2^2\right]$$

whose solution is given also in closed form by

$$\mathbf{q}_i[k+1] = \left[\mathbf{P}^\top[k+1]\breve{\boldsymbol{\mathcal{W}}}_i^\top\breve{\boldsymbol{\mathcal{W}}}_i\mathbf{P}[k+1] + \mu_{\mathbf{L}}\bar{\beta}\mathbf{I}_\rho\right]^{-1}\mathbf{P}^\top[k+1]\breve{\boldsymbol{\mathcal{W}}}_i^\top\breve{\mathbf{s}}_i \tag{3.15}$$

which again involves matrix inversion of dimension ρ-by-ρ. The overall algorithm is tabulated in Alg. 1.

Although the proposed batch algorithm exhibits low computational and memory requirements, it is not suitable for online processing, since (3.6) must be re-solved every time a new set of measurements arrive, incurring major computational burden. Thus, the development of an online recursive algorithm is well motivated.

Algorithm 1 Batch solver of (P2) in (3.6)

Initialize $\mathbf{E}[1] := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}[1]$ and $\mathbf{Q}[1]$ at random

for $k = 1, 2, \ldots$ **do**
  [S1] *Update* $\mathbf{E}$*:*
  Set $\mathbf{e} = \mathrm{vec}(\mathbf{E}[k])$
  for $l = 1, 2, \ldots, N_x N_y$ **do**
    Set $\check{\check{\mathbf{s}}}_l[k] := \check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^\top \mathrm{vec}(\mathbf{P}[k]\mathbf{Q}^\top[k]) - \boldsymbol{\mathcal{W}}_{-l}^\top \mathbf{e}_{-l}$
    Compute $e_l[k+1] = \mathrm{soft\_th}(\boldsymbol{\omega}_l^\top \check{\check{\mathbf{s}}}_l[k]; \mu_{\mathbf{E}}\bar{\beta})/\|\boldsymbol{\omega}_l\|_2^2$
  end for
  Set $\mathbf{E}[k+1] = \mathrm{unvec}(\mathbf{e}[k+1])$
  [S2] *Update* $\mathbf{P}$*:*
  for $i = 1, 2, \ldots, N_x$ **do**
    Set $\tilde{\boldsymbol{\mathcal{W}}}_i$ and $\tilde{\mathbf{s}}_i$
    Compute $\mathbf{p}_i[k+1] = \left[\mathbf{Q}^\top[k]\tilde{\boldsymbol{\mathcal{W}}}_i^\top \tilde{\boldsymbol{\mathcal{W}}}_i \mathbf{Q}[k] + \mu_{\mathbf{L}}\bar{\beta}\mathbf{I}_\rho\right]^{-1} (\mathbf{Q}^\top[k]\tilde{\boldsymbol{\mathcal{W}}}_i^\top \tilde{\mathbf{s}}_i)$
  end for
  Update $\mathbf{P}[k+1] = [\mathbf{p}_1[k+1], \mathbf{p}_2[k+1], \ldots, \mathbf{p}_{N_x}[k+1]]^\top$
  [S3] *Update* $\mathbf{Q}$*:*
  for $i = 1, 2, \ldots, N_y$ **do**
    Set $\check{\boldsymbol{\mathcal{W}}}_i$ and $\check{\mathbf{s}}_i$
    Compute $\mathbf{q}_i[k+1] = \left[\mathbf{P}^\top[k+1]\check{\boldsymbol{\mathcal{W}}}_i^\top \check{\boldsymbol{\mathcal{W}}}_i \mathbf{P}[k+1] + \mu_{\mathbf{L}}\bar{\beta}\mathbf{I}_\rho\right]^{-1} \mathbf{P}^\top[k+1]\check{\boldsymbol{\mathcal{W}}}_i^\top \check{\mathbf{s}}_i$
  end for
  Update $\mathbf{Q}[k+1] = [\mathbf{q}_1[k+1], \mathbf{q}_2[k+1], \ldots, \mathbf{q}_{N_y}[k+1]]^\top$
end for
Set $\widehat{\mathbf{P}} := \mathbf{P}[k+1]$, $\widehat{\mathbf{Q}} := \mathbf{Q}[k+1]$, and $\widehat{\mathbf{E}} := \mathbf{E}[k+1]$
return $\widehat{\mathbf{P}}$, $\widehat{\mathbf{Q}}$, and $\widehat{\mathbf{E}}$

3.2 Online Algorithm

3.2.1 Stochastic approximation approach

In practice, it is often the case that a new set of data becomes available sequentially in time. Then, it is desirable to have an algorithm that can process the newly acquired data incrementally and refine the previous estimates, rather than re-computing the batch solution, which may incur prohibitively growing computational burden. Furthermore, when the channel is time-varying due to, e.g., mobile obstacles, online algorithms can readily track such variations.

Stochastic approximation (SA) is an appealing strategy for deriving online algorithms [84, 48]. Recently, techniques involving minimizing majorized surrogate functions were developed to handle nonconvex cost functions in online settings [58, 60, 56, 80]. An online algorithm to solve a dictionary learning problem was proposed in [56]. A stochastic gradient descent algorithm was derived for subspace tracking and anomaly detection in [58]. Here, an online algorithm for the CPCP problem is developed. The proposed approach employs quadratic surrogate functions with diagonal weighting so as to capture disparate curvatures in the directions of different block variables.

For simplicity, let the number of measurements per time slot t be constant $M := |\mathcal{M}^{(t)}|$ for all t. Define $\mathbf{X} := (\mathbf{P}, \mathbf{Q}, \mathbf{E}) \in \mathcal{X} \subset \mathcal{X}' := \mathbb{R}^{(N_x \times \rho)} \times \mathbb{R}^{(N_y \times \rho)} \times \mathbb{R}^{(N_x \times N_y)}$, where $\mathcal{X}$ is a compact convex set, and $\mathcal{X}'$ a bounded open set, and $\boldsymbol{\xi}^{(t)} := [\{\check{s}_m^{(t)}\}_{m=1}^M, \{\mathbf{W}_m^{(t)}\}_{m=1}^M] \in \Xi$, where Ξ is assumed to be bounded. Define with slight abuse of notation

$$g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) := \frac{1}{2} \sum_{m=1}^{M} \left(\langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^\top + \mathbf{E} \rangle - \check{s}_m^{(t)} \right)^2 \tag{3.16}$$

$$g_2(\mathbf{X}) = g_2(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \frac{\mu_{\mathbf{L}}}{2} \left(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2 \right) + \mu_{\mathbf{E}} \|\mathbf{E}\|_1. \tag{3.17}$$

A quadratic surrogate function for $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ is then constructed as

$$\begin{aligned} \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) := g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \\ + \langle \mathbf{P} - \mathbf{P}^{(t-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{P}}^{(t)}}{2} \|\mathbf{P} - \mathbf{P}^{(t-1)}\|_F^2 \\ + \langle \mathbf{Q} - \mathbf{Q}^{(t-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{Q}}^{(t)}}{2} \|\mathbf{Q} - \mathbf{Q}^{(t-1)}\|_F^2 \\ + \langle \mathbf{E} - \mathbf{E}^{(t-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{\eta_{\mathbf{E}}^{(t)}}{2} \|\mathbf{E} - \mathbf{E}^{(t-1)}\|_F^2 \end{aligned} \tag{3.18}$$

where $\eta_{\mathbf{P}}^{(t)}$, $\eta_{\mathbf{Q}}^{(t)}$, and $\eta_{\mathbf{E}}^{(t)}$ are positive constants, and with $\tilde{\tilde{f}}_m^{(t)}(\mathbf{P}, \mathbf{Q}, \mathbf{E}) := \langle \mathbf{W}_m^{(t)}, \mathbf{P}\mathbf{Q}^\top + \mathbf{E} \rangle - \check{s}_m^{(t)}$ it can be readily verified that

$$\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \tag{3.19}$$

$$\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)\top} \mathbf{P}^{(t-1)} \tag{3.20}$$

$$\nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \sum_{m=1}^{M} \tilde{\tilde{f}}_m^{(t)}(\mathbf{P}^{(t-1)}, \mathbf{Q}^{(t-1)}, \mathbf{E}^{(t-1)}) \mathbf{W}_m^{(t)}. \tag{3.21}$$

Let us focus on the case without the forgetting factor, i.e., $\beta = 1$. A convergent SA algorithm for (P2) is obtained by considering the following surrogate problem

$$\text{(P3)} \quad \min_{\mathbf{X}} \frac{1}{t} \sum_{\tau=1}^{t} \left[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right]. \tag{3.22}$$

In fact, solving (P3) yields a stochastic gradient descent (SGD) algorithm. In particular, since variables $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$ can be separately optimized in (P3), the proposed algorithm updates the variables in parallel in each time slot t as

$$\mathbf{P}^{(t)} = \arg\min_{\mathbf{P}} \sum_{\tau=1}^{t} \left[\langle \mathbf{P} - \mathbf{P}^{(\tau-1)}, \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{P}}^{(\tau)}}{2} \|\mathbf{P} - \mathbf{P}^{(\tau-1)}\|_F^2 + \frac{\mu_{\mathbf{L}}}{2} \|\mathbf{P}\|_F^2 \right] \tag{3.23}$$

$$\mathbf{Q}^{(t)} = \arg\min_{\mathbf{Q}} \sum_{\tau=1}^{t} \left[\langle \mathbf{Q} - \mathbf{Q}^{(\tau-1)}, \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{Q}}^{(\tau)}}{2} \|\mathbf{Q} - \mathbf{Q}^{(\tau-1)}\|_F^2 + \frac{\mu_{\mathbf{L}}}{2} \|\mathbf{Q}\|_F^2 \right] \tag{3.24}$$

$$\mathbf{E}^{(t)} = \arg\min_{\mathbf{E}} \sum_{\tau=1}^{t} \left[\langle \mathbf{E} - \mathbf{E}^{(\tau-1)}, \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \rangle + \frac{\eta_{\mathbf{E}}^{(\tau)}}{2} \|\mathbf{E} - \mathbf{E}^{(\tau-1)}\|_F^2 + \mu_{\mathbf{E}} \|\mathbf{E}\|_1 \right]. \tag{3.25}$$

By checking the first-order optimality conditions, and defining $\bar{\eta}_{\mathbf{P}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{P}}^{(\tau)}$ and $\bar{\eta}_{\mathbf{Q}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{Q}}^{(\tau)}$, the update rules for $\mathbf{P}$ and $\mathbf{Q}$ are obtained as

$$\mathbf{P}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \mu_{\mathbf{L}} t} \sum_{\tau=1}^{t} \left[\eta_{\mathbf{P}}^{(\tau)} \mathbf{P}^{(\tau-1)} - \nabla_{\mathbf{P}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \right] \tag{3.26}$$

$$\mathbf{Q}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \mu_{\mathbf{L}} t} \sum_{\tau=1}^{t} \left[\eta_{\mathbf{Q}}^{(\tau)} \mathbf{Q}^{(\tau-1)} - \nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \right] \tag{3.27}$$

which can be written in recursive forms as

$$\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \mu_{\mathbf{L}} t} \left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \mu_{\mathbf{L}} \mathbf{P}^{(t-1)} \right) \tag{3.28}$$

$$\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \mu_{\mathbf{L}} t} \left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \mu_{\mathbf{L}} \mathbf{Q}^{(t-1)} \right). \tag{3.29}$$

Due to the non-smoothness of $\|\mathbf{E}\|_1$, the update for $\mathbf{E}$ proceeds in two steps. First, an auxiliary variable $\mathbf{Z}^{(t)}$ is introduced, which is computed as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\sum_{k=1}^{t} \eta_{\mathbf{E}}^{(k)} \mathbf{E}^{(k-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(k-1)}, \boldsymbol{\xi}^{(k)}) \right]. \tag{3.30}$$

Again defining $\bar{\eta}_{\mathbf{E}}^{(t)} := \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$, matrix $\mathbf{Z}^{(t)}$ can be obtained recursively as

$$\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]. \tag{3.31}$$

Then, $\mathbf{E}^{(t)}$ is updated as

$$\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu_{\mathbf{E}} t / \bar{\eta}_{\mathbf{E}}^{(t)}). \tag{3.32}$$

The overall online algorithm is tabulated in Alg. 2.

Remark 1.1 (Computational complexity). For the batch algorithm in Alg. 1, the complexity orders for computing the updates for each of $\mathbf{p}_i$ and $\mathbf{q}_i$ are $\mathcal{O}(N_y MT)$ and $\mathcal{O}(N_x MT)$, respectively, due to the computation of $\tilde{\boldsymbol{\mathcal{W}}}^{\top} \tilde{\mathbf{s}}_i$ and $\check{\boldsymbol{\mathcal{W}}}_i^{\top} \check{\mathbf{s}}_i$. Thus, the complexity orders for updating $\mathbf{P}$ and $\mathbf{Q}$ per iteration k are both $\mathcal{O}(N_x N_y MT)$. The update of e_l incurs complexity $\mathcal{O}(MT)$ for computing $\boldsymbol{\omega}_l^{\top} \check{\mathbf{s}}_l$. Thus, the complexity order for updating $\mathbf{E}$ per iteration k is $\mathcal{O}(N_x N_y MT)$. Accordingly, the overall per-iteration complexity of the batch algorithm becomes $\mathcal{O}(N_x N_y MT)$. On the other hand, the complexity of the online algorithm in Alg. 2 is dominated by the gradient computations, which require $\mathcal{O}(\rho N_x N_y M)$. Since ρ is smaller than N_x and N_y, and the per-iteration complexity does not grow with T, the online algorithm has a much more affordable complexity than its batch counterpart, and it is scalable for large network scenarios.

Algorithm 2 Online SGD solver of (P2) in (3.6)

Initialize $\mathbf{E}^{(0)} := \mathbf{0}_{N_x \times N_y}$, $\mathbf{P}^{(0)}$ and $\mathbf{Q}^{(0)}$ at random

1: **for** $t = 1, 2, \ldots$ **do**
2: Set $L_{\mathbf{P}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \mathbf{Q}^{(t-1)} \right\|_F^2$, $L_{\mathbf{Q}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)\top} \mathbf{P}^{(t-1)} \right\|_F^2$
3: Set $L_{\mathbf{E}} = \sum_{m=1}^{M} \left\| \mathbf{W}_m^{(t)} \right\|_F^2$ and $L_{\min} = \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$
4: Set $\eta_{\mathbf{P}}^{(t)} \geq \frac{L_{\mathbf{P}}^2}{L_{\min}}, \eta_{\mathbf{Q}}^{(t)} \geq \frac{L_{\mathbf{Q}}^2}{L_{\min}}$, and $\eta_{\mathbf{E}}^{(t)} \geq \frac{L_{\mathbf{E}}^2}{L_{\min}}$
5: Set $\bar{\eta}_{\mathbf{P}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{P}}^{(\tau)}$, $\bar{\eta}_{\mathbf{Q}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{Q}}^{(\tau)}$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = \sum_{\tau=1}^{t} \eta_{\mathbf{E}}^{(\tau)}$
6: Update $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{P}}^{(t)} + \mu_{\mathbf{L}} t} \left(\nabla_{\mathbf{P}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \mu_{\mathbf{L}} \mathbf{P}^{(t-1)} \right)$
7: Update $\mathbf{Q}^{(t)} = \mathbf{Q}^{(t-1)} - \frac{1}{\bar{\eta}_{\mathbf{Q}}^{(t)} + \mu_{\mathbf{L}} t} \left(\nabla_{\mathbf{Q}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \mu_{\mathbf{L}} \mathbf{Q}^{(t-1)} \right)$
8: Update $\mathbf{Z}^{(t)} = \frac{1}{\bar{\eta}_{\mathbf{E}}^{(t)}} \left[\eta_{\mathbf{E}}^{(t)} \mathbf{E}^{(t-1)} + \bar{\eta}_{\mathbf{E}}^{(t-1)} \mathbf{Z}^{(t-1)} - \nabla_{\mathbf{E}} g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \right]$
9: Set $\mathbf{E}^{(t)} = \text{soft\_th}(\mathbf{Z}^{(t)}; \mu_{\mathbf{E}} t / \bar{\eta}_{\mathbf{E}}^{(t)})$
10: **end for**

3.2.2 Convergence

The iterates $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ generated from Alg. 2 converge to a stationary point of (P2), as asserted in the following proposition. First define

$$C_t(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} \left[g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right] \tag{3.33}$$

$$\check{C}_t(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} \left[\check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) + g_2(\mathbf{X}) \right] \tag{3.34}$$

$$C(\mathbf{X}) := \mathbb{E}_{\boldsymbol{\xi}} \left[g_1(\mathbf{X}, \boldsymbol{\xi}) + g_2(\mathbf{X}) \right]. \tag{3.35}$$

Note that $C_t(\mathbf{X})$ is essentially identical to the cost of (P2). Furthermore, the minimizer of $C_t(\mathbf{X})$ approaches that of $C(\mathbf{X})$ when $t \to \infty$, provided $\boldsymbol{\xi}$ obeys the law of large numbers, which is clearly the case when e.g., $\{\boldsymbol{\xi}^{(t)}\}$ is i.i.d.

Assume that $\nabla_{\mathbf{P}} g_1(\cdot, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi})$, $\nabla_{\mathbf{Q}}(\mathbf{P}, \cdot, \mathbf{E}, \boldsymbol{\xi})$ and $\nabla_{\mathbf{E}}(\mathbf{P}, \mathbf{Q}, \cdot, \boldsymbol{\xi})$ are Lipschitz with respect to $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$, respectively, with constants $L_{\mathbf{P}}$, $L_{\mathbf{Q}}$, and $L_{\mathbf{E}}$, respectively (which will be shown in Appendix A.2). Furthermore, let $\bar{\alpha}_i^{(t)} := (\sum_{\tau=1}^{t} (\eta_i^{(\tau)} + \mu_{\mathbf{L}}))^{-1}$ for $i \in \{\mathbf{P}, \mathbf{Q}\}$, and $\bar{\alpha}_{\mathbf{E}}^{(t)} := (\bar{\eta}_{\mathbf{E}}^{(t)})^{-1}$ denote step sizes.

Proposition 2: *If* (a1) $\{\boldsymbol{\xi}^{(t)}\}_{t=1}^{\infty}$ *is an independent and identically distributed (i.i.d)*

random sequence; (a2) $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *are in a compact set* $\mathcal{X}$*;* (a3) Ξ *is bounded;* (a4) *For* $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$, $\bar{\eta}_i^{(t)} \geq ct$ $\forall t$ *for some* $c \geq 0$*; and* (a5) $c' \geq \eta_i^{(t)} \geq L_i^2/L_{\min}$ $\forall t$ *for some* $c' > 0$ *and* $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$*, then the iterates* $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ *generated by Alg. 2 converge to the set of stationary points of* (P2) *with* $\beta = 1$*, i.e.,*

$$\lim_{t\to\infty} \inf_{\bar{\mathbf{X}}\in\bar{\mathcal{X}}} \|\mathbf{X}^{(t)} - \bar{\mathbf{X}}\|_F = 0 \qquad \text{a.s.} \tag{3.36}$$

where $\bar{\mathcal{X}}$ *is the set of stationary points of* $C(\mathbf{X})$*.*
Proof: See Appendix A.2.

3.3 Numerical Tests

Performance of the proposed batch and online algorithms was assessed through numerical tests using both synthetic and real datasets. A few existing methods were also tested for comparison. The ridge-regularized least-squares (LS) scheme estimates the SLF as $\hat{\boldsymbol{f}} = (\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^\top + \omega\mathbf{C}_f^{-1})^{-1}\boldsymbol{\mathcal{W}}\check{\mathbf{s}}$, where $\mathbf{C}_f$ is the spatial covariance matrix of the SLF, and ω is a regularization parameter [89, 44, 33]. The total variation (TV)-regularized LS scheme in [73] was also tested, which solves $\min_{\boldsymbol{f}} \|\check{\mathbf{s}} - \boldsymbol{\mathcal{W}}^\top \boldsymbol{f}\|_2^2 + \omega \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y} |F_{i+1,j} - F_{i,j}| + \sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1} |F_{i,j+1} - F_{i,j}|$ where $F_{i,j} := [\mathbf{F}]_{i,j}$. Finally, the LASSO estimator was obtained by solving (P1) with $\mu_{\mathbf{L}} = 0$.

3.3.1 Test with synthetic data

Random tomographic measurements were taken by sensors deployed uniformly over $\mathcal{A} := [0.5, 40.5] \times [0.5, 40.5]$, from which the SLF with $N_x = N_y = 40$ was reconstructed. Per-time slot, 10 measurements were taken, corrupted by zero-mean white Gaussian noise with variance $\sigma^2 = 0.1$. The regularization parameters were set to $\mu_{\mathbf{L}} = 0.05$ and $\mu_{\mathbf{E}} = 0.01$ through cross-validation by minimizing the normalized error $\|\widehat{\mathbf{F}} - \mathbf{F}_0\|_F/\|\mathbf{F}_0\|_F$, where $\mathbf{F}_0$ is the ground-truth SLF depicted in Fig. 3.1. Other parameters were set to $\rho = 13$, $\beta = 1$, and $\lambda = 0.06$; while $\mathbf{C}_f = \mathbf{I}_{N_xN_y}$ and $\omega = 0.13$ were used for the ridge-regularized LS.

To validate the batch algorithm in Alg. 1, two cases were tested. In the first case, the measurements were generated for $T = 130$ time slots using $N = 52$ sensors, while

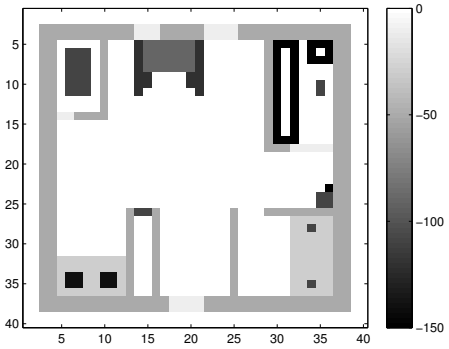
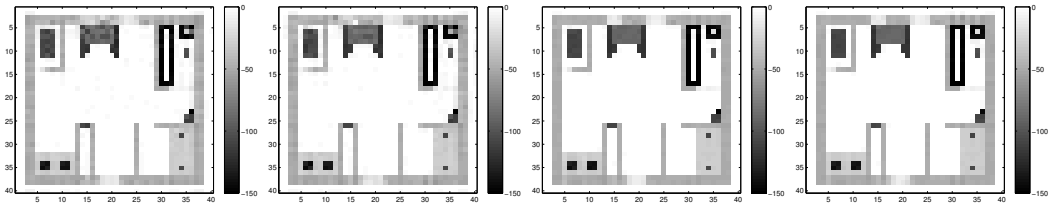


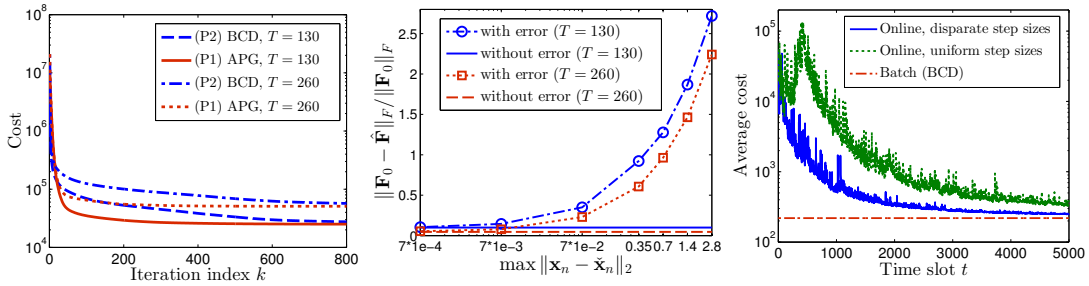
Figure 3.1: True SLF.



(a) (b) (c) (d)

Figure 3.2: Reconstructed SLFs $\widehat{\mathbf{F}}$ via batch algorithms: (a) BCD ($T = 130$, $N = 52$); (b) APG ($T = 130$, $N = 52$); (c) BCD ($T = 260$, $N = 73$); and (d) APG ($T = 260$, $N = 73$).

in the second case, $T = 260$ and $N = 73$ were used. As a comparison, the accelerated proximal gradient (APG) algorithm was also derived for (P1) [53]. Note that the APG requires the costly SVD operation of an N_x-by-N_y matrix per iteration, while only the inversion of a ρ-by-ρ matrix is necessary in the proposed BCD algorithm. Fig. 3.2 shows the SLFs reconstructed by APG and BCD algorithms for the two cases. Apparently, the reconstructed SLFs capture well the features of the ground-truth SLF in Fig. 3.1. Note that (P2) is underdetermined when $T = 130$ since the total number of unknowns in (P2) is $2,640$ while the total number of measurements is only $1,300$. This verifies that the channel gain maps can be accurately interpolated with a small number of measurements by leveraging the attributes of the low rank and sparsity. Fig. 3.3a shows the convergence of the BCD and APG algorithms. The cost of (P2) from the BCD algorithm converges to that of (P1) from APG after $k = 550$ iterations, showing that the performance of solving (P1) directly is achievable by the proposed algorithm solving



(a) (b) (c)

Figure 3.3: SLF reconstruction using the batch and online algorithms. (a) Cost versus iterations (batch). (b) Reconstruction error versus CR location error (batch). (c) Average cost over time slots (online).

Table 3.1: Reconstruction error at $T = 130$ and computational complexity per iteration.

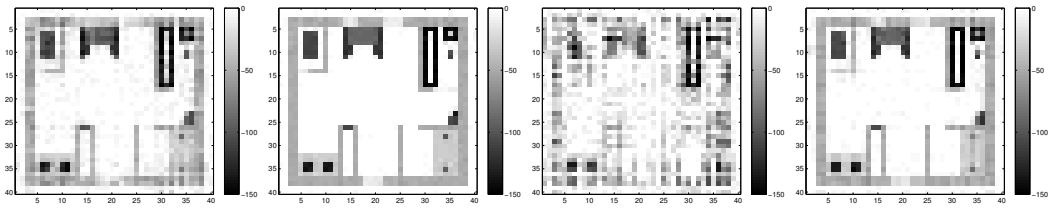
Algorithm	Proposed (BCD)	Ridge-reg. LS	TV-reg. (ADMM)	LASSO
$\|\mathbf{F}_0 - \widehat{\mathbf{F}}\|_F / \|\mathbf{F}_0\|_F$	0.1064	0.1796	0.1196	0.1828
Per-iteratoin Complexity	$\mathcal{O}(N_x N_y MT)$	N/A	$\mathcal{O}\left((N_x N_y)^3 + (N_x N_y)^2 MT\right)$	$\mathcal{O}(N_x N_y MT)$

(P2) instead. This can also be corroborated from the reconstructed SLFs in Fig. 3.2 as well.

Table 3.1 lists the reconstruction error when $T = 130$ and the per-iteration complexity of the batch algorithms. It is seen that the proposed method outperforms benchmark algorithms in terms of the reconstruction error. Note that the ridge-regularized LS has a one-shot (non-iterative) complexity of $\mathcal{O}((N_x N_y)^3)$, but its reconstruction capability is worse than the proposed algorithm as the true SLF is not smooth.

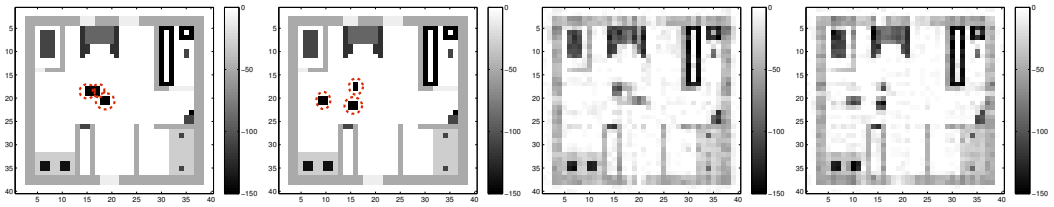
To test robustness of the proposed algorithm against imprecise CR location estimates, the reconstruction error versus the maximum sensor location error is depicted in Fig. 3.3b. To reconstruct $\mathbf{F}$, $\boldsymbol{\mathcal{W}}$ was computed via a set of erroneous sensor locations $\check{\mathbf{x}}_n^{(t)}$ obtained by adding uniformly random perturbations to true locations $\mathbf{x}_n^{(t)}$. It is seen that the SLF could be accurately reconstructed when the location error was small.

The numerical tests for the online algorithm were carried out with the same parameter setting as the batch experiments with $N = 317$. Fig. 3.3c depicts the evolution of the average cost in (3.33) for two sets of values for $(\bar{\eta}_{\mathbf{P}}^{(t)}, \bar{\eta}_{\mathbf{Q}}^{(t)}, \bar{\eta}_{\mathbf{E}}^{(t)})$. The green dotted curve corresponds to using $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$, while the blue solid curve is for $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$, and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$. It can be seen that the uniform step sizes for all



(a) $t = 1,000$ (b) $t = 5,000$ (c) $t = 1,000$ (d) $t = 5,000$

Figure 3.4: Reconstructed SLFs $\widehat{\mathbf{F}}$ by the online algorithm with (a)-(b) $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 300$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 10$; and (c)-(d) $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = \bar{\eta}_{\mathbf{E}}^{(t)} = 300$.



(a) $t = 2,400$ (b) $t = 3,200$ (c) $t = 2,400$ (d) $t = 3,200$

Figure 3.5: (a)-(b) True SLFs $\mathbf{F}_0^{(t)}$ and (c)-(d) reconstructed SLFs $\widehat{\mathbf{F}}^{(t)}$ at different time slots.

variables result in convergence rate that is slower than that with the disparate step sizes. Fig. 3.4 shows the SLFs reconstructed via the online algorithm at $t = 1,000$ and $t = 5,000$ using the two choices of step sizes. It can be seen that for a given time slot t, flexibly choosing the step sizes yields much more accurate reconstruction. As far as reconstruction error, the online algorithm with disparate step sizes yields 6.3×10^{-2} at $t = 5,000$, while its batch counterpart has 2.4×10^{-2}. Although slightly less accurate SLF is obtained by the online algorithm, it comes with greater computational efficiency.

To assess the tracking ability of the online algorithm, the slow channel variation was simulated. The measurements were generated using the SLF in Fig. 3.1 with three additional objects slowly moving in the rate of unit pixel width per 70 time slots. Fig. 3.5 depicts instances of the true and reconstructed SLFs at $t = 2,400$ and $t = 3,200$, respectively, obtained by the online algorithm. The moving objects are marked by the red circles. It is seen that the reconstructed SLFs correctly capture the moving objects, while the stationary objects are estimated more clearly as t increases.

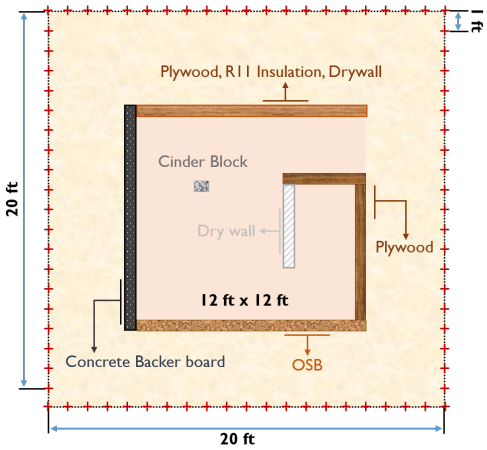
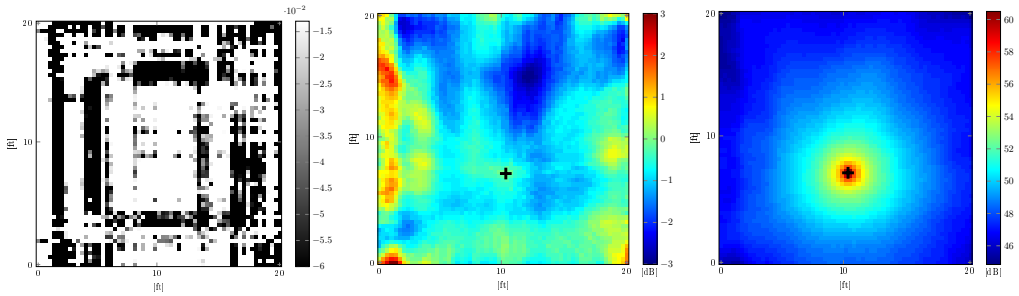


Figure 3.6: Configuration of the testbed with $N = 80$ sensor locations marked with crosses.

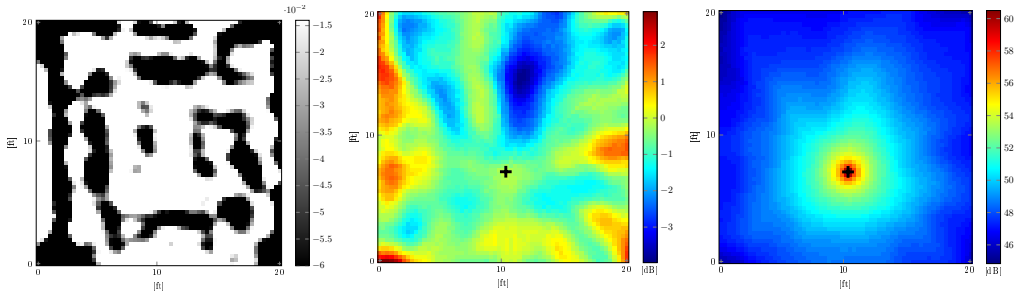


(a) SLF (b) Shadow fading map (c) Channel gain map

Figure 3.7: Reconstructions by the proposed batch algorithm in Alg. 1.

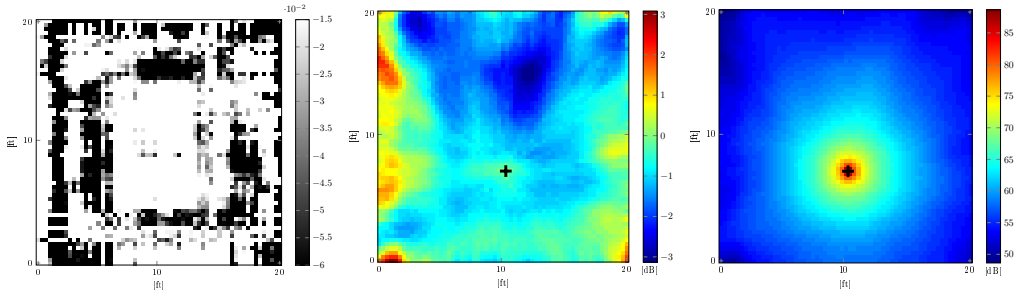
3.3.2 Test with real data

To validate the performance of the proposed framework for SLF and channel gain map estimation in realistic scenarios, real received signal strength (RSS) measurements were also processed. The data were collected by a set of $N = 80$ sensors deployed in the perimeter of a square-shaped testbed as shown in Fig. 3.6, where the crosses indicate the sensor positions. Data collection was performed in two steps [33]. First, free-space measurements were taken to obtain estimates of the path gain g_0 and the pathloss exponent γ via least-squares. The estimated γ was approximately 2, and g_0 was found to be 75. Then, tomographic measurements were formed with the artificial structure shown



(a) SLF (b) Shadow fading map (c) Channel gain map

Figure 3.8: Reconstructions by the ridge-regularized LS.



(a) SLF (b) Shadow fading map (c) Channel gain map

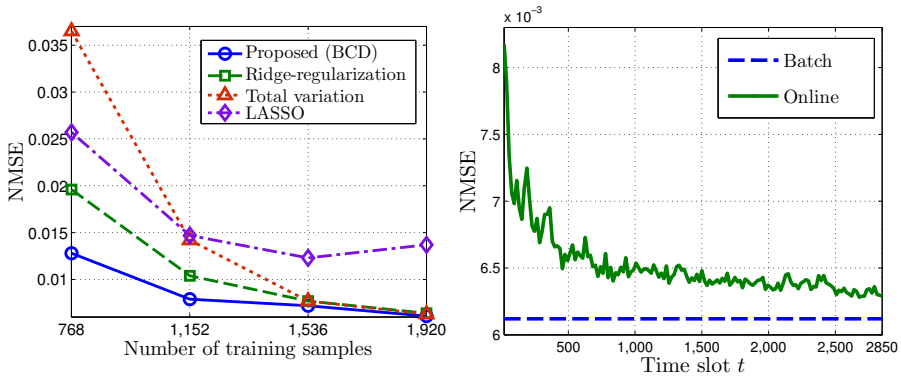
Figure 3.9: Reconstructions by the proposed online algorithm in Alg. 2.

in Fig. 3.6. For the both measurements, 100 measurements were taken per time slot, in the 2.425 GHz frequency band, across 24 time slots. The shadowing measurements were obtained by subtracting the estimated pathloss from the RSS measurements.

The SLFs of size $N_x = N_y = 61$ were reconstructed by the proposed batch algorithm. The regularization parameters were set to $\mu_{\mathbf{L}} = 4.5$ and $\mu_{\mathbf{E}} = 3.44$, which were determined by cross-validation. The parameter λ in (2.3) was set to 0.2 to capture the non-zero weights within the first Fresnel zone, and $\rho = 10$ and $\beta = 1$ were used.

For comparison, the ridge-regularized LS estimator was also tested. To construct $\mathbf{C}_f$, the exponential decay model in [2] was used, which models the covariance between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ as $\left[\mathbf{C}_f\right]_{ij} = \sigma_s^2 \exp[-\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2/\kappa]$, where σ_s^2 and $\kappa > 0$ are model parameters. In our tests, $\sigma_s^2 = \kappa = 1$, and $\omega = 79.9$ were used.

The SLF, shadow fading map, and channel gain map reconstructed by the proposed BCD algorithm are depicted in Fig. 3.7. The shadow fading and channel gain maps



(a) (b)

Figure 3.10: NMSE of channel gain prediction by (a) the batch; and (b) online algorithms.

portray the gains in dB between any point in the map and the fixed CR location at $(10.2, 7.2)$ (marked by the cross). Fig. 3.8 shows the results from the ridge-regularized LS estimation. It can be seen from Fig. 3.7a and Fig. 3.8a that the proposed low-rank plus sparse model produces a somewhat sharper SLF image than the ridge-regularized LS approach. Although the latter yields a smooth SLF image, it produces more artifacts near the isolated block and the boundary of the SLF. Such artifacts may lead to less accurate shadowing and channel gain maps. For instance, Fig. 3.7b and Fig. 3.8b both show that the shadow fading is stronger as more building material is crossed in the communication path. However, somewhat strong attenuations are observed near the cinder block location and the interior of the oriented strand board (OSB) wall only in Fig. 3.8b, which seems anomalous.

The online algorithm was also tested with the real data. Parameters $\bar{\eta}_{\mathbf{P}}^{(t)} = \bar{\eta}_{\mathbf{Q}}^{(t)} = 620$ and $\bar{\eta}_{\mathbf{E}}^{(t)} = 200$ were selected, and 6×10^5 measurements were uniformly drawn from the original dataset with replacement to demonstrate the asymptotic performance. Fig. 3.9 depicts the reconstructed SLF, shadow fading and channel gain maps obtained from the online algorithm. It can be seen that the SLF shown in Fig. 3.9a is close to that depicted in Fig. 3.7a. Similar observations can be made for the shadow fading and channel gain maps as well. Thus, the online algorithm is a viable alternative to the batch algorithm with reduced computational complexity, and affordable memory requirement.

Channel gain estimation performance of the proposed algorithms was assessed via 5-fold cross-validation. Let $\check{\boldsymbol{g}}_{\text{test}}$ and $\widehat{\boldsymbol{g}}_{\text{test}}$ denote RSS measurement vectors in the test set and its estimate, respectively. Prediction performance is measured by the normalized mean-square error (NMSE) $\|\check{\boldsymbol{g}}_{\text{test}} - \widehat{\boldsymbol{g}}_{\text{test}}\|^2/\|\check{\boldsymbol{g}}_{\text{test}}\|^2$. Fig. 3.10a displays the NMSE of batch algorithms with 480 test samples versus the number of training samples. It is shown that the proposed algorithm outperforms competing alternatives, particularly when a small number of training samples are available, validating the usefulness of the proposed model. The online algorithm was also tested with 2.85×10^5 measurements uniformly drawn from $1,920$ training samples with replacement. Fig. 3.10b depicts the evolution of the NMSE measured on 480 test samples at every t. It is observed that the online algorithm attains the batch performance as t increases.

3.4 Conclusion

A low-rank plus sparse matrix model was proposed for channel-gain cartography, which is instrumental for various CR spectrum sensing and resource allocation tasks. The channel gain was modeled as the sum of the distance-based path-loss and the tomographic accumulation of the underlying SLF for shadowing. The SLF was postulated to have a low-rank structure corrupted by sparse outliers. Efficient batch and online algorithms were derived by leveraging a bifactor-based characterization of the matrix nuclear norm. The algorithms enjoy low computational complexity and a reduced memory requirement, without sacrificing the optimality, with provable convergence properties. Tests with both synthetic and real measurement datasets corroborated the claims and showed that the algorithms could accurately reveal the structure of the propagation medium.

Chapter 4

Bayesian Approach to Channel-gain Cartography

4.1 Motivation

Conventionally, the spatial loss field (SLF) $\boldsymbol{f}$ is learned via regularized least-squares (LS) methods tailored to the propagation environment [33, 51, 89], by solving

$$\min_{\boldsymbol{f}} \sum_{\tau=1}^{t} \left(\check{s}_\tau - \sum_{i=1}^{N_g} w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_i) f(\tilde{\mathbf{x}}_i) \right)^2 + \mu_f \mathcal{R}(\boldsymbol{f}) \tag{4.1}$$

where $\mathcal{R} : \mathbb{R}^{N_g} \to \mathbb{R}$ is a generic regularizer to promote a known attribute of $\boldsymbol{f}$, and $\mu_f \geq 0$ is a regularization scalar to reflect compliance of $\boldsymbol{f}$ with this attribute. Particularly, a ridge-regularized solution can be interpreted as a maximum a posteriori (MAP) estimator provided that the SLF is statistically homogeneous and modeled as a zero-mean Gaussian random field. However, these approaches are less effective when the propagation environment is spatially heterogeneous due to a combination of free space and objects in different sizes and materials (e.g., as easily seen in urban areas), which subsequently induces statistical heterogeneity in the SLF. To account for environmental heterogeneity, the novel method here leverages the Bayesian framework to learn the piecewise homogeneous SLF through a hidden Markov random field (MRF) model [38], which captures spatial correlations of neighboring regions exhibiting related statistical

behavior. Efficient field estimators will be derived by using Markov chain Monte Carlo (MCMC) sampling [30], which is a powerful tool for Bayesian inference when analytical solutions of the minimum mean-square error (MMSE) or the MAP estimators are not available. Furthermore, hyperparameters are estimated as well, instead of being fixed a priori.

Besides accounting for heterogeneous propagation, another contribution here is a data-adaptive sensor selection technique, with the goal of reducing SLF uncertainty, by cross-fertilizing ideas from the fields of experimental design [26] and active learning [55]. The conditional entropy of the SLF is considered as an uncertainty measure in this work, giving rise to a novel sensor selection criterion. Although such criterion is intractable especially when the size of the SLF is large, its efficient proxy can be obtained thanks to the availability of posterior samples from the proposed MCMC-based algorithm. Note that the proposed technique is appealing for a practical scenario constrained to incur low communication overhead, since the data collection cost can be reduced by using a minimal number of selective measurements to learn the SLF.

4.2 Adaptive Bayesian Channel-gain Cartography

In this section, we view $\boldsymbol{f}$ as random, and forth propose a two-layer Bayesian SLF model, along with an MCMC-based approach for inference. We further develop an adaptive data acquisition strategy to select informative measurements.

4.2.1 Bayesian Model and Problem Formulation

Let $\mathcal{A}$ consist of two disjoint homogeneous regions $\mathcal{A}_0 := \{\mathbf{x}|\mathbb{E}[f(\mathbf{x})] = \mu_{f_0}, \mathrm{Var}[f(\mathbf{x})] = \sigma^2_{f_0}, \mathbf{x} \in \mathcal{A}\}$, and $\mathcal{A}_1 := \{\mathbf{x}|\mathbb{E}[f(\mathbf{x})] = \mu_{f_1}, \mathrm{Var}[f(\mathbf{x})] = \sigma^2_{f_1}, \mathbf{x} \in \mathcal{A}\}$, giving rise to a hidden label field $\boldsymbol{z} := [z(\tilde{\mathbf{x}}_1), \ldots, z(\tilde{\mathbf{x}}_{N_g})]^\top \in \{0,1\}^{N_g}$ with binary entries $z(\tilde{\mathbf{x}}_i) = k$ if $\tilde{\mathbf{x}}_i \in \mathcal{A}_k \ \forall i$, and $k = 0, 1$. The two separate regions can be used to model heterogeneous environments. For instance, if $\mathcal{A}$ corresponds to an urban area, $\mathcal{A}_1$ may include densely populated regions with buildings, while $\mathcal{A}_0$ with $\mu_{f_0} < \mu_{f_1}$ may capture the less obstructive open spaces. In such a paradigm, we model the conditional distribution of $f(\tilde{\mathbf{x}}_i)$ as

$$f(\tilde{\mathbf{x}}_i)|z(\tilde{\mathbf{x}}_i) = k \sim \mathcal{N}(\mu_{f_k}, \sigma^2_{f_k}), \tag{4.2}$$

while the Ising prior [85], which is a binary version of the discrete MRF Potts prior [38], is assigned to $\boldsymbol{z}$ in order to capture the dependency among spatially correlated labels. By the Hammersley-Clifford theorem [34], the Ising prior of $\boldsymbol{z}$ follows a Gibbs distribution

$$p(\boldsymbol{z}|\beta) = \frac{1}{C(\beta)} \exp\left[\beta \sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i))\right] \tag{4.3}$$

where $\mathcal{N}(\tilde{\mathbf{x}}_i)$ is a set of indices associated with 1-hop neighbors of $\tilde{\mathbf{x}}_i$ on the rectangular grid in Fig. 4.1, β is a granularity coefficient controlling the degree of homogeneity in z, $\delta(\cdot)$ is Kronecker's delta, and

$$C(\beta) := \sum_{\boldsymbol{z} \in \mathcal{Z}} \exp\left[\beta \sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i))\right] \tag{4.4}$$

is the partition function with $\mathcal{Z} := \{0,1\}^{N_g}$. By assuming conditional independence of $\{f(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g}$ given $\boldsymbol{z}$, the resulting model is referred to as the Gauss-Markov-Potts model with two labels. The Gauss-Markov-Potts model for channel-gain cartography is depicted in Fig. 4.2 with the measurement model in (2.2).

To describe priors of other parameters, let ν_t be independent and identically distributed (i.i.d.) Gaussian with zero mean and variance σ_ν^2, and $\boldsymbol{\theta}$ denote a hyperparameter vector comprising σ_ν^2, β, and $\boldsymbol{\theta}_f := [\mu_{f_0}, \mu_{f_1}, \sigma_{f_0}^2, \sigma_{f_1}^2]^\top$. The weight matrix $\mathbf{W}_t \in \mathbb{R}^{N_g \times t}$ is formed with columns $\mathbf{w}_\tau^{(n,n')} := [w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_1), \ldots, w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_{N_g})]^\top := [w_{\tau,1}, \ldots, w_{\tau,N_g}]^\top \in \mathbb{R}^{N_g}$ of the link $\mathbf{x}_{n(\tau)}$–$\mathbf{x}_{n'(\tau)}$ for $\tau = 1, \ldots, t$. Assuming the independence among entries of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ can be expressed as

$$p(\boldsymbol{\theta}) = p(\sigma_\nu^2)p(\beta)p(\boldsymbol{\mu}_{f_k})p(\boldsymbol{\sigma}_{f_k}^2) \tag{4.5}$$

with $p(\boldsymbol{\mu}_{f_k}) = p(\mu_{f_0})p(\mu_{f_1})$ and $p(\boldsymbol{\sigma}_{f_k}^2) = p(\sigma_{f_0}^2)p(\sigma_{f_1}^2)$, where the individual priors $p(\sigma_\nu^2)$, $p(\beta)$, $p(\boldsymbol{\mu}_{f_k})$, and $p(\boldsymbol{\sigma}_{f_k}^2)$ are specified next.

1) Granularity coefficient β. To cope with the variability of β in accordance with structural patterns of the propagation medium, β is viewed as an unknown random variable that is to be estimated together with $\boldsymbol{f}$ and $\boldsymbol{z}$ under the Bayesian framework.

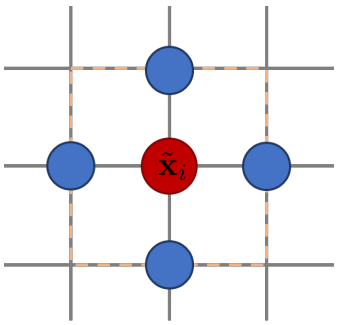


Figure 4.1: Four-connected MRF with $z(\tilde{\mathbf{x}}_i)$ marked red and its neighbors in $\mathcal{N}(\tilde{\mathbf{x}}_i)$ marked blue.

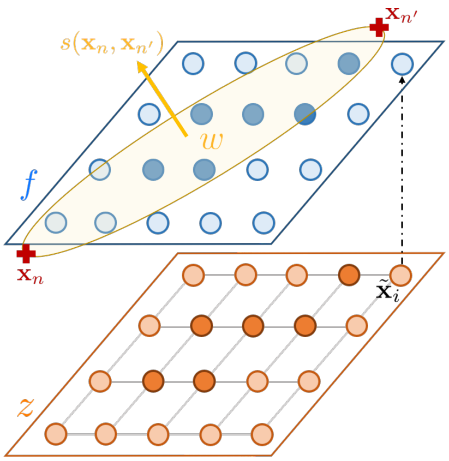


Figure 4.2: The Gauss-Markov-Potts model with Ising prior for channel-gain cartography, together with the measurement model for sensors located at $(\mathbf{x}_n, \mathbf{x}_{n'})$.

Similar to e.g., [72], the uniform distribution is adopted for the prior of β as

$$p(\beta) = \mathcal{U}_{(0,\beta_{\max})}(\beta) := \begin{cases} 1/\beta_{\max}, & \text{if } \beta \in [0, \beta_{\max}] \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

2) Noise variance σ_ν^2. In the presence of the additive Gaussian noise with fixed mean, it is common to assign a conjugate prior to σ_ν^2, which reproduces a posterior distribution in the same family of its prior. The inverse gamma (IG) distribution serves this purpose

for $\sigma_\nu^2 \in \mathbb{R}^+$ as follows:

$$p(\sigma_\nu^2) = \mathcal{IG}(a_\nu, b_\nu) := \frac{b_\nu^{a_\nu}}{\Gamma(a_\nu)}(\sigma_\nu^2)^{-a_\nu-1}\exp\left(-\frac{b_\nu}{\sigma_\nu^2}\right) \tag{4.7}$$

where a_ν is referred to as the shape parameter, b_ν as the scale parameter, and $\Gamma(\cdot)$ denotes the gamma function.

3) Hyperparameters of the SLF $\boldsymbol{\theta}_f$. While the prior for μ_{f_k} is assumed to be Gaussian with mean m_k and variance $\sigma_k^2 \in \mathbb{R}^+$ (see also [5]), the inverse Gamma distribution parameterized by $\{a_k, b_k\}$ is considered for the prior of $\sigma_{f_k}^2$:

$$p(\mu_{f_k}) = \mathcal{N}(m_k, \sigma_k^2),\ k = 0, 1 \tag{4.8}$$

$$p(\sigma_{f_k}^2) = \mathcal{IG}(a_k, b_k),\ k = 0, 1. \tag{4.9}$$

Such choice of the conjugate priors in (4.8) and (4.9) provides analytical tractability for estimating μ_{f_k}. Note that a truncated Gaussian prior for μ_{f_k} also can be adopted when the support of μ_{f_k} is known a priori.

Together with the priors for $\{\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}\}$, our joint posterior becomes

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\check{\mathbf{s}}_t|\boldsymbol{f}, \sigma_\nu^2)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{z}|\beta)p(\boldsymbol{\theta}) \tag{4.10}$$

where $p(\check{\mathbf{s}}_t|\boldsymbol{f}, \sigma_\nu^2) \sim \mathcal{N}(\mathbf{W}_t^\top \boldsymbol{f}, \sigma_\nu^2\mathbf{I}_t)$ is the data likelihood with the weight matrix $\mathbf{W}_t \in \mathbb{R}^{N_g \times t}$ formed from columns $\mathbf{w}_\tau^{(n,n')} := [w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_1), \ldots, w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_{N_g})]^\top := [w_{\tau,1}, \ldots, w_{\tau,N_g}]^\top \in \mathbb{R}^{N_g}$ of the link $\mathbf{x}_{n(\tau)}$–$\mathbf{x}_{n'(\tau)}$ for $\tau = 1, \ldots, t$. Note that Fig. 4.3 summarizes the proposed hierarchical Bayesian model for $\{\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}\}$ as a directed acyclic graph, where the dependency between (hyper) parameters is indicated with an arrow.

We will pursue the the conditional MMSE estimator

$$\hat{\boldsymbol{f}}_{\text{MMSE}} := \mathbb{E}[\boldsymbol{f}|\boldsymbol{z} = \hat{\boldsymbol{z}}_{\text{MAP}}, \check{\mathbf{s}}_t] \tag{4.11}$$

where the marginal MAP estimate is

$$\hat{\boldsymbol{z}}_{\text{MAP}} := \arg\max_{\boldsymbol{z}} p(\boldsymbol{z}|\check{\mathbf{s}}_t). \tag{4.12}$$

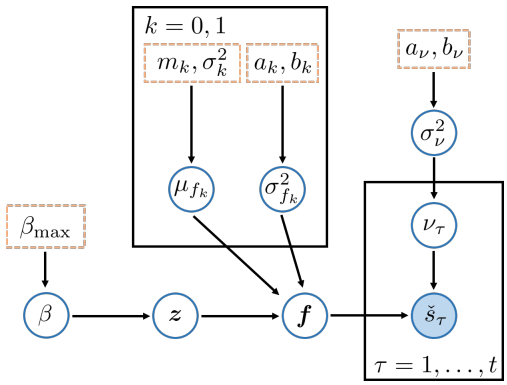


Figure 4.3: Graphical representation of the hierarchical Bayesian model with Ising prior for (hyper) parameters (those in boxes are fixed).

Furthermore, the marginal MMSE estimates of the $\boldsymbol{\theta}$ entries are found as

$$\widehat{\sigma_\nu^2}_{\text{MMSE}} := \mathbb{E}[\sigma_\nu^2|\check{\mathbf{s}}_t] \tag{4.13}$$

$$\widehat{\beta}_{\text{MMSE}} := \mathbb{E}[\beta|\check{\mathbf{s}}_t] \tag{4.14}$$

$$\widehat{\mu_{f_k}}_{\text{MMSE}} := \mathbb{E}[\mu_{f_k}|\check{\mathbf{s}}_t], \ k = 0, 1 \tag{4.15}$$

$$\widehat{\sigma_{f_k}^2}_{\text{MMSE}} := \mathbb{E}[\sigma_{f_k}^2|\check{\mathbf{s}}_t], \ k = 0, 1. \tag{4.16}$$

4.2.2 Approximate Inference via Markov Chain Monte Carlo

While approximate estimators have been proposed for Bayesian inference (see e.g., [42, 98]), analytical solutions to (4.11)−(4.16) are not tractable due to the complex form of the posterior in (4.10) that does not permit marginalization or maximization. To bypass this challenge, one can generate samples from (4.10), and then numerically approximate the desired estimators from those samples. MCMC is a class of methods used to generate samples from a complex distribution [30].

Among MCMC methods, Gibbs sampling [29] is particularly suitable for this work. It draws samples following the target distribution (e.g., the posterior in (4.10)) by sweeping through each variable to sample from its conditional distribution while fixing the others to their up-to-date values. Although the samples at early iterations of Gibbs sampling with random initialization are not representative of the desired distribution (such duration is called the *burn-in* period $N_{\text{Burn-in}}$), the theory of MCMC guarantees that

Algorithm 3 Metropolis-within-Gibbs sampler for $\{\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}\}$

Input: $\boldsymbol{z}^{(0)}$, $\boldsymbol{\theta}^{(0)}$, $\check{\mathbf{s}}_t$, $\mathbf{W}_t$, N_{CL}, $N_{\text{Burn-in}}$, and N_{Iter}

1: **for** $l = 1$ to N_{Iter} **do**
2: Generate $\boldsymbol{f}^{(l)} \sim p(\boldsymbol{f}|\check{\mathbf{s}}_t, \boldsymbol{z}^{(l-1)}, \boldsymbol{\theta}^{(l-1)})$ in (4.17)
3: Generate $\boldsymbol{z}^{(l)} \sim p(\boldsymbol{z}|\check{\mathbf{s}}_t, \boldsymbol{f}^{(l)}, \boldsymbol{\theta}^{(l-1)})$ via Alg. 4
4: Generate $\beta^{(l)} \sim p(\beta|\check{\mathbf{s}}_t, \boldsymbol{f}^{(l)}, \boldsymbol{z}^{(l)}, \sigma_\nu^{2(l-1)}, \boldsymbol{\theta}_f^{(l-1)})$ via Alg. 6
5: Generate $\sigma_\nu^{2(l)} \sim p(\sigma_\nu^2|\check{\mathbf{s}}_t, \boldsymbol{f}^{(l)}, \boldsymbol{z}^{(l)}, \beta^{(l)}, \boldsymbol{\theta}_f^{(l-1)})$ in (4.27)
6: Generate $\mu_{f_k}^{(l)} \sim p(\mu_{f_k}|\check{\mathbf{s}}_t, \boldsymbol{f}^{(l)}, \boldsymbol{z}^{(l)}, \sigma_\nu^{2(l)}, \beta^{(l)}, \boldsymbol{\sigma}_{f_k}^{2(l-1)})$ in (4.29) for $k = 0, 1$
7: Generate $\sigma_{f_k}^{2(l)} \sim p(\sigma_{f_k}^2|\check{\mathbf{s}}_t, \boldsymbol{f}^{(l)}, \boldsymbol{z}^{(l)}, \sigma_\nu^{2(l)}, \beta^{(l)}, \boldsymbol{\mu}_{f_k}^{(l)})$ in (4.33) for $k = 0, 1$
8: **end for**
9: **return** $\mathcal{S}^{(t)} := \left\{\boldsymbol{f}^{(l)}, \boldsymbol{z}^{(l)}, \boldsymbol{\theta}^{(l)}\right\}_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}}$

the stationary distribution of those samples matches with the target distribution [30].

Gibbs sampling requires only the conditional distribution within a proportionality scale. When a given conditional distribution is not easy to simulate, one can resort to a Metropolis-Hastings (MH) sampler [36], which generates a candidate from a simple proposal distribution of such conditional distribution, and accepts (or rejects) the candidate as a sample of interest under a certain acceptance ratio α. The substitution of MH sampling for some sampling steps inside the Gibbs sampler results in a Metropolis-within-Gibbs (MwG) sampler, as listed in Alg. 3. Posterior conditionals considered in this work and associated sampling methods will be described next.

1) Spatial loss field $\boldsymbol{f}$. It is easy to show that

$$\begin{aligned} p(\boldsymbol{f}|\check{\mathbf{s}}_t, \boldsymbol{z}, \boldsymbol{\theta}) &\propto p(\check{\mathbf{s}}_t|\boldsymbol{f}, \sigma_\nu^2) p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f) \\ &= \mathcal{N}(\check{\boldsymbol{\mu}}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}, \boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}) \end{aligned} \tag{4.17}$$

where

$$\boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t} := \left((\sigma_\nu^2)^{-1}\mathbf{W}_t\mathbf{W}_t^\top + \boldsymbol{\Sigma}_{f|z}^{-1}\right)^{-1} \tag{4.18}$$

$$\check{\boldsymbol{\mu}}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t} := \boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\left((\sigma_\nu^2)^{-1}\mathbf{W}_t\check{\mathbf{s}}_t + \boldsymbol{\Sigma}_{f|z}^{-1}\boldsymbol{\mu}_{f|z}\right) \tag{4.19}$$

since $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)$ follows $\mathcal{N}(\boldsymbol{\mu}_{f|z}, \boldsymbol{\Sigma}_{f|z})$ by (4.2), with $\boldsymbol{\Sigma}_{f|z} := \text{diag}(\{\text{Var}[f_i|z_i]\}_{i=1}^{N_g})$ and $\boldsymbol{\mu}_{f|z} := \mathbb{E}[\boldsymbol{f}|\boldsymbol{z}]$ where $f_i := f(\tilde{\mathbf{x}}_i)$ and $z_i := z(\tilde{\mathbf{x}}_i)$ (see Appendix B.1 for derivation).

Algorithm 4 Single-site Gibbs sampler for $\boldsymbol{z}$

Input: $\boldsymbol{f}^{(l)}$ and $\boldsymbol{z}^{(l-1)}$

Initialize $\boldsymbol{\zeta}^{(l)} := [\zeta_1^{(l)}, \ldots, \zeta_{N_g}^{(l)}]^\top = \boldsymbol{z}^{(l-1)}$
for $i = 1$ to N_g **do**
  Obtain h_i in (4.21) with $\boldsymbol{z} = \boldsymbol{\zeta}^{(l)}$ and $\boldsymbol{f} = \boldsymbol{f}^{(l)}$
  Generate $u \sim \mathcal{U}_{(0,1)}$
  if $u < (1 + h_i)^{-1}$ **then**
    Set $\zeta_i^{(l)} = 1$
  else
    Set $\zeta_i^{(l)} = 0$
  end if
end for
return $\boldsymbol{z}^{(l)} = \boldsymbol{\zeta}^{(l)}$

Hence, $\boldsymbol{f}$ can be easily simulated by a standard sampling method.

2) Hidden label field $\boldsymbol{z}$. A Gibbs sampler is required to simulate $p(\boldsymbol{z}|\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{\theta}) \propto p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{z}|\beta)$ while avoiding the intractable computation of $C(\beta)$ in (4.4). Let $\boldsymbol{z}_{-i}$ and $\boldsymbol{z}_{\mathcal{N}(\tilde{\mathbf{x}}_i)}$ represent replicas of $\boldsymbol{z}$ without its i-th entry, and only with the entries of $\mathcal{N}(\tilde{\mathbf{x}}_i)$, respectively. By the Markovianity of $\boldsymbol{z}$ and conditional independence between f_i and f_j $\forall i \neq j$ given $\boldsymbol{z}$, the conditional distribution of z_i is

$$p(z_i|\boldsymbol{z}_{-i}, \check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{\theta}) \propto \exp\left[\ell(z_i) + \beta \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z_j - z_i)\right] \tag{4.20}$$

where $\ell(z_i) := \ln p(f_i|z_i, \boldsymbol{\theta}_f)$. After evaluating (4.20) for $z_i = 0, 1$ and normalizing, one can obtain $p(z_i = 1|\boldsymbol{z}_{-i}, \check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{\theta}) = (1 + h_i)^{-1}$, where

$$h_i := \exp\left[\ell(z_i = 0) - \ell(z_i = 1) + \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta(1 - 2z_j)\right] \tag{4.21}$$

with $\delta(z_j) - \delta(z_j - 1) = 1 - 2z_j$. Then, the sample of $\boldsymbol{z}$ can be obtained via the single-site Gibbs sampler by using (4.21), as summarized in Alg. 4. It is worth stressing that the sampling criterion with h_i in (4.21) does not require the evaluation of $C(\beta)$.

3) Granularity coefficient β. The conditional distribution of β satisfies the following

proportionality relation

$$p(\beta|\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \sigma_\nu^2, \boldsymbol{\theta}_f) \propto p(\boldsymbol{z}|\beta)p(\beta)$$
$$\propto \frac{1}{\beta_{\max} C(\beta)} \exp\left[\beta \sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z_j - z_i)\right] \tag{4.22}$$

for $\beta \in [0, \beta_{\max}]$, simply by the Gibbs distribution in (4.3) and the uniform prior of β in (4.6). Unfortunately, sampling of β is formidably challenging because evaluating the partition function $C(\beta)$ in $p(\boldsymbol{z}|\beta)$, incurs exponential complexity. To address this, one may resort to auxiliary variable MCMC methods that do not require exact evaluation of $p(\boldsymbol{z}|\beta)$, including the single auxiliary variable method (SAVM) [64] and the exchange algorithm [68]. Those methods replace $C(\beta)$ with its single-point importance sampling estimate by using an auxiliary variable, which unfortunately must be generated via exact sampling that is generally expensive for statistical models with intractable partition functions. To bypass exact sampling for generating this auxiliary variable, we will leverage a *double*-MH sampling method for β; also [52].

Let $\boldsymbol{z}^*$ and β^* denote the auxiliary variable of $\boldsymbol{z}$ and a candidate of β for MH sampling, respectively. The idea behind the double-MH algorithm is to generate $\boldsymbol{z}^*$ through N_{CL} cycles of MH updates from the current sample $\boldsymbol{z}^{(l)}$, instead of using exact sampling from $p(\boldsymbol{z}^*|\beta^*)$. As the name suggests, the double-MH sampling includes two nested MCMC samplers: the inner one to generate a chain of the auxiliary variable at each step of the outer sampler for β. It is instructive to mention that N_{CL} is not necessarily large by initializing the chain with $\boldsymbol{z}^{(l)}$ at the l-th iteration [72, 52], which means that additional complexity to generate the auxiliary variable is not necessarily high. In this work, $\boldsymbol{z}^*$ is obtained via another single-site Gibbs sampler, as described in Alg. 5:

$$p(z_i^*|\boldsymbol{z}_{-i}^*, \beta^*) \propto \exp\left[\beta^* \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z_j^* - z_i^*)\right] \ \forall\ i \tag{4.23}$$

and a sample of z_i^* is generated by utilizing $p(z_i^* = 1|\boldsymbol{z}_{-i}^*, \beta^*) = (1 + h_i^*)^{-1}$ with

$$h_i^* := \exp\Big[\sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta(1 - 2z_j^*)\Big]. \tag{4.24}$$

Algorithm 5 Single-site Gibbs sampler for $\boldsymbol{z}^*$

Input: $\boldsymbol{z}^{(l)}$, β^*, and N_{CL}
Initialize $\boldsymbol{\zeta}^* := [\zeta_1^*, \ldots, \zeta_{N_g}^*]^\top = \boldsymbol{z}^{(l)}$
for $m = 1$ to N_{CL} **do**
  for $i = 1$ to N_g **do**
    Obtain h_i^* in (4.24) with $\boldsymbol{z}^* = \boldsymbol{\zeta}^*$
    Generate $u \sim \mathcal{U}_{(0,1)}$
    if $u < (1 + h_i^*)^{-1}$ **then**
      Set $\zeta_i^* = 1$
    else
      Set $\zeta_i^* = 0$
    end if
  end for
end for
return $\boldsymbol{z}^* = \boldsymbol{\zeta}^*$

The overall double-MH sampler for β is summarized in Alg. 6. A proposal distribution of β^* is the truncated Gaussian

$$q(\beta^*|\beta^{(l-1)}) = \begin{cases} \mathcal{N}(\beta^{(l-1)}, \sigma_q^2)/c, & \text{if } \beta^* \in [0, \beta_{\max}] \\ 0, & \text{otherwise} \end{cases} \tag{4.25}$$

with a tunable variable σ_q^2 and a normalizing constant

$$c := \int_0^{\beta_{\max}} \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left[-\frac{1}{2\sigma_q^2}\left(\beta^* - \beta^{(l-1)}\right)^2 \right] d\beta^*. \tag{4.26}$$

4) Noise variance σ_ν^2. With $p(\sigma_\nu^2)$ in (4.7), we have the posterior conditional of σ_ν^2 satisfying

$$\begin{aligned} p(\sigma_\nu^2|\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \beta, \boldsymbol{\theta}_f) &\propto p(\check{\mathbf{s}}_t|\boldsymbol{f}, \sigma_\nu^2) p(\sigma_\nu^2) \\ &\propto \mathcal{IG}(a_\nu + \frac{t}{2}, b_\nu + \frac{1}{2}\|\check{\mathbf{s}}_t - \mathbf{W}_t^\top \boldsymbol{f}\|_2^2). \end{aligned} \tag{4.27}$$

Therefore, a sample of σ_ν^2 can be generated by a standard sampling method.

5) Means of the SLF $\boldsymbol{\mu}_{f_k}$. Let $\boldsymbol{f}_k$ be the $N_k \times 1$ vector formed by concatenating $f(\tilde{\mathbf{x}}_i)$ for $\tilde{\mathbf{x}}_i \in \mathcal{A}_k$, for $k = 0, 1$. By recalling the priori independence between the parameters of

Algorithm 6 Double-MH sampler for β

Input: $\beta^{(l-1)}$, $\boldsymbol{z}^{(l)}$, and N_{CL}

Generate $\beta^* \sim q(\beta^*|\beta^{(l-1)})$ in (4.25)
Generate $\boldsymbol{z}^* \sim p(\boldsymbol{z}^*|\beta^*)$ via Alg. 5
Set $\alpha' := \frac{p(\beta^*)q(\beta^{(l-1)}|\beta^*)p(\boldsymbol{z}^*|\beta^{(l-1)})p(\boldsymbol{z}^{(l)}|\beta^*)}{p(\beta^{(l-1)})q(\beta^*|\beta^{(l-1)})p(\boldsymbol{z}^{(l)}|\beta^{(l-1)})p(\boldsymbol{z}^*|\beta^*)}$
Obtain $\alpha = \min\{1, \alpha'\}$
Generate $u \sim \mathcal{U}_{(0,1)}$
if $u < \alpha$ **then**
  $\beta^{(l)} = \beta^*$
else
  $\beta^{(l)} = \beta^{(l-1)}$
end if
return $\beta^{(l)}$

disjoint homogeneous regions $\mathcal{A}_0$ and $\mathcal{A}_1$, the posterior conditional of $\boldsymbol{\mu}_{f_k} := [\mu_{f_0}, \mu_{f_1}]^\top$ can be expressed as

$$\begin{aligned} p(\boldsymbol{\mu}_{f_k}|\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \sigma_\nu^2, \beta, \sigma_{f_0}^2, \sigma_{f_1}^2) &\propto p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f) p(\boldsymbol{\mu}_{f_k}) \\ &\propto p(\mu_{f_0}|\boldsymbol{z}, \boldsymbol{f}_0, \sigma_{f_0}^2) p(\mu_{f_1}|\boldsymbol{z}, \boldsymbol{f}_1, \sigma_{f_1}^2) \end{aligned} \tag{4.28}$$

with

$$p(\mu_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \sigma_{f_k}^2) \propto p(\boldsymbol{f}_k|\boldsymbol{z}, \mu_{f_k}, \sigma_{f_k}^2) p(\mu_{f_k}), \ \forall k. \tag{4.29}$$

Since a sample of each μ_{f_k} can be independently drawn according to $p(\mu_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \sigma_{f_k}^2)$ in (4.29), the sampling method for μ_{f_k} will be described.

To efficiently simulate a sample of μ_{f_k}, the likelihood $p(\boldsymbol{f}_k|\boldsymbol{z}, \mu_{f_k}, \sigma_{f_k}^2)$ is recast as an univariate distribution with respect to the sample mean $\bar{f}_k := (\sum_i f_{k,i})/N_k$ as

$$\begin{aligned} p(\boldsymbol{f}_k|\boldsymbol{z}, \mu_{f_k}, \sigma_{f_k}^2) &\propto \exp\left[-\frac{1}{2\sigma_{f_k}^2} \sum_{i=1}^{N_k} (f_{k,i} - \mu_{f_k})^2\right] \\ &\propto \exp\left[-\frac{1}{2\sigma_{f_k}^2} (-2\mu_{f_k} \sum_{i=1}^{N_k} f_{k,i} + N_k \mu_{f_k}^2)\right] \\ &\propto \exp\left[-\frac{N_k}{2\sigma_{f_k}^2} (\bar{f}_k - \mu_{f_k})^2\right] \\ &\propto \mathcal{N}(\mu_{f_k}, 2\sigma_{f_k}^2/N_k). \end{aligned} \tag{4.30}$$

Since $p(\mu_{f_k})$ is the Gaussian conjugate prior, one can show that $p(\mu_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \sigma^2_{f_k})$ is Gaussian as well, parameterized by

$$\begin{aligned}\mathbb{E}\left[\mu_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \sigma^2_{f_k}\right] &= \frac{\sigma_k^2 \bar{f}_k}{\sigma_k^2 + (\sigma^2_{f_k}/N_k)} + \frac{\sigma^2_{f_k}/N_k}{\sigma_k^2 + (\sigma^2_{f_k}/N_k)} m_k \\ \text{Var}\left[\mu_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \sigma^2_{f_k}\right] &= \left(\frac{1}{\sigma_k^2} + \frac{N_k}{\sigma^2_{f_k}}\right)^{-1}. \end{aligned} \tag{4.31}$$

Therefore, a sample of μ_{f_k} can be generated for $k = 0, 1$ by using a standard sampling method.

6) Variances of the SLF $\boldsymbol{\sigma}^2_{f_k}$. Similar to $\boldsymbol{\mu}_{f_k}$, the statistical independence between $\mathcal{A}_0$ and $\mathcal{A}_1$ leads to the following proportionality of the posterior conditional for $\boldsymbol{\sigma}^2_{f_k} := [\sigma^2_{f_0}, \sigma^2_{f_1}]^\top$

$$\begin{aligned} p(\boldsymbol{\sigma}^2_{f_k}|\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \sigma^2_\nu, \beta, \boldsymbol{\mu}_{f_k}) &\propto p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f) p(\boldsymbol{\sigma}^2_{f_k}) \\ &\propto p(\sigma^2_{f_0}|\boldsymbol{z}, \boldsymbol{f}_0, \mu_{f_0}) p(\sigma^2_{f_1}|\boldsymbol{z}, \boldsymbol{f}_1, \mu_{f_1}) \end{aligned} \tag{4.32}$$

where

$$\begin{aligned} p(\sigma^2_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \mu_{f_k}) &\propto p(\boldsymbol{f}_k|\boldsymbol{z}, \mu_{f_k}, \sigma^2_{f_k}) p(\sigma^2_{f_k}) \\ &\propto \mathcal{IG}(a_k + \frac{N_k}{2}, b_k + \frac{1}{2}\|\boldsymbol{f}_k - \mu_{f_k}\mathbf{1}_{N_k}\|_2^2), \ \forall k. \end{aligned} \tag{4.33}$$

Therefore, a sample of each σ^2_k can be independently drawn according to $p(\sigma^2_{f_k}|\boldsymbol{z}, \boldsymbol{f}_k, \mu_{f_k})$ in (4.33).

4.2.3 Efficient Sample-based Estimators

In this section, efficient sample-based estimators for $\boldsymbol{f}$, $\boldsymbol{z}$, and $\boldsymbol{\theta}$ are derived, by using a set of samples $\mathcal{S}^{(t)}$ from Alg. 3. Building on [42], the elementwise marginal MAP estimator of $\boldsymbol{z}$ and its sample-based approximation are

$$\hat{z}_{i,\text{MAP}} = \arg \max_{z_i \in \{0,1\}} p(z_i|\check{\mathbf{s}}_t)$$

$$\simeq \arg \max_{z_i \in \{0,1\}} \frac{1}{|\mathcal{S}^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \delta(z_i^{(l)} - z_i) \tag{4.34}$$

for $i = 1, \ldots, N_g$. After obtaining $\widehat{\boldsymbol{z}}_{\text{MAP}}$, the sample-based elementwise conditional MMSE estimator of $\boldsymbol{f}$ follows as

$$\hat{f}_{i,\text{MMSE}} \simeq \frac{1}{|\mathcal{S}_i^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} f_i^{(l)} \delta(z_i^{(l)} - \widehat{z}_{i,\text{MAP}}), \quad \forall i \tag{4.35}$$

where $\mathcal{S}_i^{(t)} \subset \mathcal{S}^{(t)}$ is a subset of samples such that $z_i^{(l)} = \widehat{z}_{i,\text{MAP}}$ for $l = N_{\text{Burn-in}} + 1, \ldots, N_{\text{Iter}}$. To estimate $\boldsymbol{\theta}$, the following marginal MMSE estimators are employed

$$\widehat{\beta}_{\text{MMSE}} \simeq \frac{1}{|\mathcal{S}^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \beta^{(l)} \tag{4.36}$$

$$\widehat{\sigma_\nu^2}_{\text{MMSE}} \simeq \frac{1}{|\mathcal{S}^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \sigma_\nu^{2(l)} \tag{4.37}$$

$$\widehat{\mu_{f_k}}_{\text{MMSE}} \simeq \frac{1}{|\mathcal{S}^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \mu_{f_k}^{(l)}, \ k = 0, 1 \tag{4.38}$$

$$\widehat{\sigma_{f_k}^2}_{\text{MMSE}} \simeq \frac{1}{|\mathcal{S}^{(t)}|} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \sigma_{f_k}^{2(l)}, \ k = 0, 1. \tag{4.39}$$

Remark 4.1 (Monitoring sampler-convergence). The proposed sampler in Alg. 3 generates a sequence of samples from the desired distribution in (4.10), after a burn-in period to diminish the influence of initialization. By recalling that the stationary distribution of those samples is matched with the desired distribution, monitoring convergence of sample-sequences guides the choice of $N_{\text{Burn-in}}$.

Let ψ denote a generic scalar random variable of interest. Suppose that N_{Seq} parallel sequences of length N_{Iter} are available, and let $\psi^{(l,l')}$ denote the l-th sample of ψ in the l'-th sequence for $l = 1, \ldots, N_{\text{Iter}}$ and $l' = 1, \ldots, N_{\text{Seq}}$. Then, the following potential scale reduction factor (PSRF) estimate is adopted for convergence diagnosis [30]

$$\text{PSRF}(\psi) := \frac{N'_{\text{Iter}} - 1}{N'_{\text{Iter}}} + \frac{\sigma^2_{\text{Between}}}{\sigma^2_{\text{Within}}} \tag{4.40}$$

where $N'_{\text{Iter}} := N_{\text{Iter}} - N_{\text{Burn-in}}$, the within-sequence variance:

$$\sigma^2_{\text{Within}} := \frac{1}{N_{\text{Seq}}} \sum_{l'=1}^{N_{\text{Seq}}} \frac{1}{N'_{\text{Iter}} - 1} \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \left(\psi^{(l,l')} - \bar{\psi}^{(l')}\right)^2 \tag{4.41}$$

with $\bar{\psi}^{(l')} := \sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} \psi^{(l,l')}/(N'_{\text{Iter}} - 1)\ \forall l'$, and the between-sequence variance:

$$\sigma^2_{\text{Between}} := \frac{1}{N_{\text{Seq}}} \sum_{l'=1}^{N_{\text{Seq}}} \left(\bar{\psi}^{(l')} - \bar{\psi}\right)^2 \tag{4.42}$$

with $\bar{\psi} := \sum_{l'=1}^{N_{\text{Seq}}} \bar{\psi}^{(l')}/N_{\text{Seq}}$. As those sequences converge while $N_{\text{Iter}} \to \infty$, the PSRF declines to 1. In practice, each sequence is supposed to follow the desired distribution when PSRF ≤ 1.2 [30, p. 138]. For synthetic data tests, $N_{\text{Burn-in}}$ and N_{Iter} were found to have PSRF ≤ 1.06 for $\boldsymbol{f}$, $\boldsymbol{z}$, and $\boldsymbol{\theta}$ over $N_{\text{Seq}} = 20$ independent sequences. On the other hand, $N_{\text{Burn-in}}$ and N_{Iter} for real data tests were found to have PSRF ≤ 1.04 for $\boldsymbol{f}$ and $\boldsymbol{z}$, while the PSRF < 1.5 for $\boldsymbol{\theta}$, over $N_{\text{Seq}} = 20$ independent sequences. It allows to have moderate-sized $N_{\text{Burn-in}}$ and N_{Iter} for real data tests. Note that elementwise $\{\text{PSNR}(f_i), \text{PSNR}(z_i)\}_{i=1}^{N_g}$ were monitored for $\boldsymbol{f}$ and $\boldsymbol{z}$.

Remark 4.2 (Computational complexity). For the proposed MCMC method in Alg. 3, the complexity order to generate a sample of $\boldsymbol{f}$ is $\mathcal{O}(N_g^3)$ per iteration l to compute $\boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}$ in (4.18). While sampling of $\boldsymbol{z}$ incurs complexity $\mathcal{O}(N_g)$, that of $\boldsymbol{\theta}$ has complexity $\mathcal{O}(N_g N_{\text{CL}})$ dominated by the sampling required for β via Alg. 6. Therefore, the overall computational complexity per iteration l is $\mathcal{O}(N_g^3 + N_g(N_{\text{CL}} + 1)) \approx \mathcal{O}(N_g^3)$ for $N_{\text{CL}} \ll N_g$. Note that $N_{\text{CL}} = 2$ is used for numerical tests, while $N_g \approx 1.6 \times 10^3$.

For conventional methods to estimate $\boldsymbol{f}$, the ridge regularized LS [33] has a one-shot (non-iterative) complexity of $\mathcal{O}(N_g^3)$, while the total variation (TV) regularized LS via the alternating direction method of multipliers (ADMM) in [73] incurs complexity of $\mathcal{O}(N_g^3)$ per iteration l; see also [51, 82] for details. This shows that the computational complexity per iteration of the proposed algorithm is comparable with that of the TV regularized solution that relies on the ADMM. Extra complexity is needed to decide $N_{\text{Burn-in}}$ by checking the PSRF as described in Remark 4.1, which is computed by using multiple sample-sequences generated in parallel. However, sample-sequence generation through parallel processing saves the delay from serially generating multiple sample

sequences. Furthermore, the computational burden is kept low by the data-adaptive sensor selection strategy, which will be introduced in Sec. 4.2.4, by reducing the number of measurements to reconstruct the SLF.

4.2.4 Data-adaptive Sensor Selection

The proposed Bayesian channel-gain cartography accounts for the uncertainty of $\boldsymbol{f}$, through the variance in (4.18). Using the latter, our idea is to adaptively collect a measurement (or a mini-batch of measurements) from the set of available sensing radio pairs, with the goal of reducing the uncertainty of $\boldsymbol{f}$. To this end, we will rely on the conditional entropy [17] that in our context is given by

$$H_\tau(\boldsymbol{f}|\check{\mathbf{s}}_\tau,\boldsymbol{z},\boldsymbol{\theta}) = \sum_{\boldsymbol{z}'\in\mathcal{Z}}\int_{\boldsymbol{\theta}',\check{\mathbf{s}}'_\tau} p(\check{\mathbf{s}}'_\tau,\boldsymbol{z}',\boldsymbol{\theta}') \times H_\tau(\boldsymbol{f}|\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau,\boldsymbol{z}=\boldsymbol{z}',\boldsymbol{\theta}=\boldsymbol{\theta}')d\boldsymbol{\theta}'d\check{\mathbf{s}}'_\tau \tag{4.43}$$

where

$$\begin{aligned} H_\tau(\boldsymbol{f}|\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau,\boldsymbol{z}=\boldsymbol{z}',\boldsymbol{\theta}=\boldsymbol{\theta}') &:= -\int p(\boldsymbol{f}|\check{\mathbf{s}}'_\tau,\boldsymbol{z}',\boldsymbol{\theta}')\ln p(\boldsymbol{f}|\check{\mathbf{s}}'_\tau,\boldsymbol{z}',\boldsymbol{\theta}')d\boldsymbol{f} \\ &= \frac{1}{2}\ln\Big(\big|\boldsymbol{\Sigma}_{f|z',\boldsymbol{\theta}',\check{\mathbf{s}}'_t}\big|\Big) + \frac{N_g}{2}\Big(1+\ln(2\pi)\Big) \end{aligned} \tag{4.44}$$

and $|\cdot|$ denotes matrix determinant. To obtain $\check{s}_{\tau+1}$, one can choose a pair of sensors (n^*,n'^*), for which $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ minimizes $H_{\tau+1}(\boldsymbol{f}|\check{\mathbf{s}}_{\tau+1},\boldsymbol{z},\boldsymbol{\theta})$. Given $\check{\mathbf{s}}_\tau$, we write

$$\begin{aligned} H_{\tau+1}(\boldsymbol{f}|\check{\mathbf{s}}_{\tau+1},\boldsymbol{z},\boldsymbol{\theta}) &= H_\tau(\boldsymbol{f}|\check{\mathbf{s}}_\tau,\boldsymbol{z},\boldsymbol{\theta}) \\ &\quad - \sum_{\boldsymbol{z}'\in\mathcal{Z}}\int_{\boldsymbol{\theta}',\check{\mathbf{s}}'_{\tau+1}} p(\check{\mathbf{s}}'_{\tau+1},\boldsymbol{z}',\boldsymbol{\theta}')h(\boldsymbol{z}',\boldsymbol{\theta}',\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)d\boldsymbol{\theta}'d\check{\mathbf{s}}'_{\tau+1} \end{aligned} \tag{4.45}$$

with $h(\boldsymbol{z},\boldsymbol{\theta},\mathbf{w},\check{\mathbf{s}}_t) := \ln\left(1+(\sigma_\nu^2)^{-1}\mathbf{w}^\top\boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\mathbf{w}\right)/2$, and seek $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ by solving

$$\text{(P1)} \quad \max_{\substack{\mathbf{w}_{\tau+1}^{(n,n')}:\\ (n,n')\in\mathcal{M}_{\tau+1}}} \mathbb{E}_{\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_\tau}[h(\boldsymbol{z},\boldsymbol{\theta},\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)] = \sum_{\boldsymbol{z}'\in\mathcal{Z}}\int_{\boldsymbol{\theta}'} p(\boldsymbol{z}',\boldsymbol{\theta}'|\check{\mathbf{s}}_\tau)h(\boldsymbol{z}',\boldsymbol{\theta}',\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)d\boldsymbol{\theta}' \tag{4.46}$$

where $\mathcal{M}_\tau := \{(n,n')|\exists(\mathbf{x}_n\text{–}\mathbf{x}_{n'}) \text{ at } \tau, (n,n') \in \{1,\ldots,N\}^2\}$ is a set of available sensing radio pairs at time slot τ (see Appendix A.1 for derivation of (P1)). Note that solving (P1) in (4.46) to find $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ does not require $p(\boldsymbol{z}', \boldsymbol{\theta}'|\check{\mathbf{s}}_{\tau+1})$, which means the joint posterior in (4.10) does not need to be retrained for adaptive data acquisition.

Apparently, solving (P1) is not an easy task as evaluating $\mathbb{E}_{\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_\tau}[h(\boldsymbol{z},\boldsymbol{\theta},\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)]$ is intractable especially for large N_g since $|\mathcal{Z}| = 2^{N_g}$. Fortunately, the samples from Alg. 3 can be used to approximate

$$\mathbb{E}_{\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_\tau}[h(\boldsymbol{z},\boldsymbol{\theta},\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)] \simeq \frac{1}{|\mathcal{S}^{(\tau)}|}\sum_{l=N_{\text{Burn-in}}+1}^{N_{\text{Iter}}} h(\boldsymbol{z}^{(l)},\boldsymbol{\theta}^{(l)},\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau) =: \bar{h}(\mathbf{w}_{\tau+1}^{(n,n')}). \tag{4.47}$$

Therefore, $\check{s}_{\tau+1}$ can be obtained from the pair of sensors corresponding to $\mathbf{w}_{\tau+1}^{(n,n')}$ with the maximum value of $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (4.47).

The steps involved for adaptive Bayesian channel-gain cartography are listed in Alg. 7.

Remark 4.3 (Mini-batch setup). The proposed adaptive sensor selection method can be easily extended to a mini-batch setup of size N_{Batch} per time slot τ as follows: i) find weight vectors $\{\mathbf{w}_{\tau+1}^{(n^{(m)},n'^{(m)})}\}_{m=1}^{N_{\text{Batch}}}$ for $\{(n^{(m)},n'^{(m)})\}_{m=1}^{N_{\text{Batch}}} \subset \mathcal{M}_{\tau+1}$ associated with N_{Batch} largest values of $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (4.47), and collect the corresponding measurements $\{\check{s}_{\tau+1}^{(m)}\}_{m=1}^{N_{\text{Batch}}}$ (steps 7–8 in Alg. 7); and ii) aggregate those measurements below $\check{\mathbf{s}}_\tau$ to construct $\check{\mathbf{s}}_{\tau+1} := [\check{\mathbf{s}}_\tau^\top, \check{s}_{\tau+1}^{(1)},\ldots,\check{s}_{\tau+1}^{(N_{\text{Batch}})}]^\top$ (step 9 in Alg. 7). Numerical tests will be performed to assess the mini-batch operation of Alg. 7.

4.3 Numerical Tests

Performance of the proposed algorithms was assessed through numerical tests using both synthetic and real datasets. A few existing methods were also tested for comparison, including the ridge-regularized SLF estimate given by $\widehat{\boldsymbol{f}}_{\text{LS}} = (\mathbf{W}_t\mathbf{W}_t^\top + \mu_f\mathbf{C}_f^{-1})^{-1}\mathbf{W}_t\check{\mathbf{s}}_t$ [33], where $\mathbf{C}_f$ is a spatial covariance matrix modeling the similarity between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ in area $\mathcal{A}$. We further tested the total variation (TV)-regularized LS scheme in [73],

Algorithm 7 Adaptive Bayesian channel-gain cartography

Input: $\boldsymbol{z}^{(0)}$, $\boldsymbol{\theta}^{(0)}$, $\check{\mathbf{s}}^{(0)}$, $\mathbf{W}^{(0)}$, g_0, γ, N_{CL}, $N_{\text{Burn-in}}$, and N_{Iter}.

1: Set $\check{\mathbf{s}}_0 = \check{\mathbf{s}}^{(0)}$ and $\mathbf{W}_0 = \mathbf{W}^{(0)}$
2: **for** $\tau = 0, 1, \ldots$ **do**
3: Obtain $\mathcal{S}^{(\tau)}$ via Alg. 3$(\boldsymbol{z}^{(0)}, \boldsymbol{\theta}^{(0)}, \check{\mathbf{s}}_\tau, \mathbf{W}_\tau, N_{\text{CL}}, N_{\text{Burn-in}}, N_{\text{Iter}})$
4: Obtain $\hat{\boldsymbol{z}}^{(\tau)}_{\text{MAP}}$ from (4.34) by using $\mathcal{S}^{(\tau)}$
5: Obtain $\hat{\boldsymbol{f}}^{(\tau)}_{\text{MMSE}}$ from (4.35) by using $\hat{\boldsymbol{z}}^{(\tau)}_{\text{MAP}}$ and $\mathcal{S}^{(\tau)}$
6: Obtain $\hat{\boldsymbol{\theta}}^{(\tau)}_{\text{MMSE}}$ from (4.36)–(4.39) by using $\mathcal{S}^{(\tau)}$
7: Evaluate $\bar{h}(\mathbf{w}^{(n,n')}_{\tau+1})$ in (4.47) $\forall (n, n') \in \mathcal{M}_{\tau+1}$ by using $\mathcal{S}^{(\tau)}$
8: Collect $\check{s}_{\tau+1}$ from (n^*, n'^*) associated with $\bar{h}(\mathbf{w}^{(n,n')}_{\tau+1})$
9: Construct $\check{\mathbf{s}}_{\tau+1} = [\check{\mathbf{s}}_\tau^\top, \check{s}_{\tau+1}]^\top$ and $\mathbf{W}_{\tau+1} := [\mathbf{W}_\tau, \mathbf{w}^{(n^*,n'^*)}_{\tau+1}]$
10: Set $\boldsymbol{z}^{(0)} = \hat{\boldsymbol{z}}^{(\tau)}_{\text{MAP}}$ and $\boldsymbol{\theta}^{(0)} = \hat{\boldsymbol{\theta}}^{(\tau)}_{\text{MMSE}}$
11: **end for**
12: Specify arbitrary locations of interest $\{\mathbf{x}, \mathbf{x}'\} \in \mathcal{A}$
13: Estimate $\hat{s}(\mathbf{x}, \mathbf{x}')$ via (2.4) by using $\hat{\boldsymbol{f}}_{\text{MMSE}}$
14: Estimate $\hat{g}(\mathbf{x}, \mathbf{x}')$ via (2.1) by using g_0, γ, and $\hat{s}(\mathbf{x}, \mathbf{x}')$

which solves the regularized problem in (4.1) with

$$\mathcal{R}(\boldsymbol{f}) = \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} |F_{i+1,j} - F_{i,j}| + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} |F_{i,j+1} - F_{i,j}|, \tag{4.48}$$

where $\mathbf{F} := \text{unvec}(\boldsymbol{f}) \in \mathbb{R}^{N_x \times N_y}$ and $F_{i,j} := [\mathbf{F}]_{i,j}$. As a competing alternative of the proposed adaptive sampling, simple *random sampling* was considered for both regularized LS estimators, by selecting $\{\check{s}^{(m)}_{\tau+1}\}_{m=1}^{N_{\text{Batch}}}$ $\forall \tau$ uniformly at random. Particularly, Alg. 7 after replacing steps 7–8 with random sampling is named as the *non-adaptive Bayesian algorithm*, and will be compared with the proposed method throughout synthetic and real data tests.

4.3.1 Test with synthetic data

This section validates the proposed algorithm through synthetic tests. Random shadowing measurements were taken by $N = 120$ sensors uniformly deployed on boundaries of $\mathcal{A} := [0.5, 40.5] \times [0.5, 40.5]$, from which the SLF defined over a grid $\{\tilde{\mathbf{x}}_i\}_{i=1}^{1,600} := \{1, \ldots, 40\}^2$ was reconstructed. To generate the ground-truth SLF $\boldsymbol{f}_0$, the hidden label

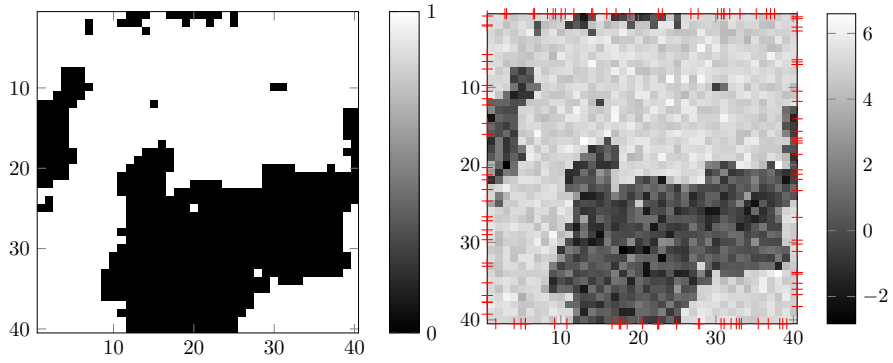
Table 4.1: Hyper-hyperparameters of $\boldsymbol{\theta}$ for synthetic tests.

$\beta_{\max}$	m_0	m_1	$\sigma_k^2, \forall k$	a_ν	b_ν	$a_k, \forall k$	$b_k, \forall k$
10	0.5	4.5	0.1	0.1	0.1	0.1	0.1

field $\boldsymbol{z}_0$ was obtained first via the Metropolis algorithm [62] by using the prior of $\boldsymbol{z}$ in (4.3) with $\beta = 1.3$. Afterwards, $\boldsymbol{f}_0$ was constructed to have $f(\tilde{\mathbf{x}}_i) \sim \mathcal{N}(0.2, 1)\ \forall \tilde{\mathbf{x}}_i \in \mathcal{A}_0$ and $f(\tilde{\mathbf{x}}_j) \sim \mathcal{N}(5, 0.2)\ \forall \tilde{\mathbf{x}}_j \in \mathcal{A}_1$ resulting in $\boldsymbol{\theta}_f = [0.2, 5, 1, 0.2]^\top$, respectively, based on labels in $\boldsymbol{z}_0$. True $\mathbf{F}_0 := \text{unvec}(\boldsymbol{f}_0) \in \mathbb{R}^{40\times 40}$ and $\mathbf{Z}_0 := \text{unvec}(\boldsymbol{z}_0) \in \{0,1\}^{40\times 40}$ are depicted in Fig. 4.4 with sensor locations marked with crosses. The effects of calibration are not accounted for this section, meaning that g_0 and γ are assumed to be known, and the fusion center directly uses shadowing measurements $\check{\mathbf{s}}_\tau$. Under the mini-batch operation, each measurement $\check{s}_\tau^{(m)}\ \forall \tau, m$ was generated according to (2.5), where s_τ was obtained by (2.4) with w set to the normalized ellipse model in (2.3), while ν_τ was set to follow zero-mean Gaussian with $\sigma_\nu^2 = 5 \times 10^{-2}$. To construct $\mathcal{M}_{\tau+1}$ per time slot τ, $|\mathcal{M}_{\tau+1}| = 100$ pairs of sensors were uniformly selected at random with replacement. Then, $N_{\text{Batch}} = 40$ shadowing measurements were collected from $\left\{ \left(n^{(m)}, n'^{(m)}\right) \right\}_{m=1}^{N_{\text{Batch}}} \subset \mathcal{M}_{\tau+1}$ to execute Alg. 7 for $\tau = 0, \ldots, 15$.

In all synthetic tests, the following simulation parameters were used: $N_{\text{CL}} = 2$, $N_{\text{Burn-in}} = 200$, $N_{\text{Iter}} = 500$, and $\sigma_q^2 = 0.03$ were used to run the proposed algorithm; and hyper-hyperparameters of $\boldsymbol{\theta}$ were set as listed in Table 4.1. For initialization, $\boldsymbol{\theta}^{(0)}$ was set to have $\beta^{(0)} = 0.1$, $\boldsymbol{\mu}_{f_k}^{(0)} = [m_0, m_1]^\top$, and randomly initialized σ_ν^2 and $\boldsymbol{\sigma}_{f_k}^2$. Vector $\boldsymbol{z}^{(0)}$ was obtained by drawing $z_i^{(0)} \sim \text{Bern}(0.5)$ for $i = 1, \ldots, N_g$, where Bern(0.5) denotes the Bernoulli distribution with mean equal to 0.5. Furthermore, $\check{\mathbf{s}}^{(0)}$ was collected from randomly selected 100 pairs of sensors. To find μ_f of the competing alternatives, the L-curve [49, Chapter 26] was used for the ridge regularization, while the generalized cross-validation [31] was adopted for the TV regularization.

The first experiment was performed to validate the efficacy of Alg. 7. The estimates $\widehat{\mathbf{F}} = \text{unvec}(\hat{\boldsymbol{f}})$ and $\widehat{\mathbf{Z}} = \text{unvec}(\hat{\boldsymbol{z}})$ at $\tau = 15$ are displayed in Figs. 4.5c and 4.5d, respectively, together with the estimated SLFs from the regularized-LS estimators in Figs. 4.5a and 4.5b. The most satisfactory result was obtained by the proposed method since piecewise homogeneous regions of the SLF were separately reconstructed by introducing the hidden label field.



(a) (b)

Figure 4.4: True fields for synthetic tests: (a) hidden label field $\mathbf{Z}_0$ and (b) spatial loss field $\mathbf{F}_0$ with $N = 120$ sensor locations marked with crosses.

To test the proposed adaptive sensor selection method, $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{Z}}$ reconstructed by the non-adaptive Bayesian algorithm are shown in Figs. 4.5e and 4.5f, respectively. Comparison between Figs. 4.5c and 4.5e visually demonstrates that improved SLF reconstruction performance could be achieved through adaptive data acquisition with the same number of measurements. Accuracy of $\hat{\boldsymbol{z}}$ was also quantitatively measured by the labeling-error, defined as $\|\boldsymbol{z}_0 - \hat{\boldsymbol{z}}\|_1/N_g$. Fig. 4.6 displays the progression of the labeling-error averaged over 20 independent Monte Carlo runs. It shows that the proposed adaptive method consistently outperforms the non-adaptive one, which implies that selection of informative measurements to decrease uncertainty of $\boldsymbol{f}$ given current estimates of $\boldsymbol{z}$ and $\boldsymbol{\theta}$ could lead to more accurate estimates of $\boldsymbol{f}$ and $\boldsymbol{z}$ in the next time slot. Meanwhile, average estimates of $\boldsymbol{\theta}$ and associated standard deviation denoted with $\pm$ are listed in Table 4.2, where every hyperparameter was accurately estimated. Together with the result in Fig. 4.5, the accurate estimates of the hyperparameters confirm that the proposed method can faithfully capture patterns of objects in area of interest, and also reveal the underlying statistical properties.

The next experiment tests robustness of the proposed algorithms against measurement noise ν_τ. The normalized error $\|\boldsymbol{f}_0 - \hat{\boldsymbol{f}}\|_2/\|\boldsymbol{f}_0\|_2$ and the labeling-error for $\boldsymbol{z}$ averaged over sensor locations and realizations of $\{\nu_\tau\}_{\tau=1}^t$ were used to quantify the reconstruction performance. Fig. 4.7 depicts the progression of those errors as a function of σ_ν^2 averaged over 20 Monte Carlo runs. Note that Figs. 4.5c–4.5e and 4.5d–4.5f

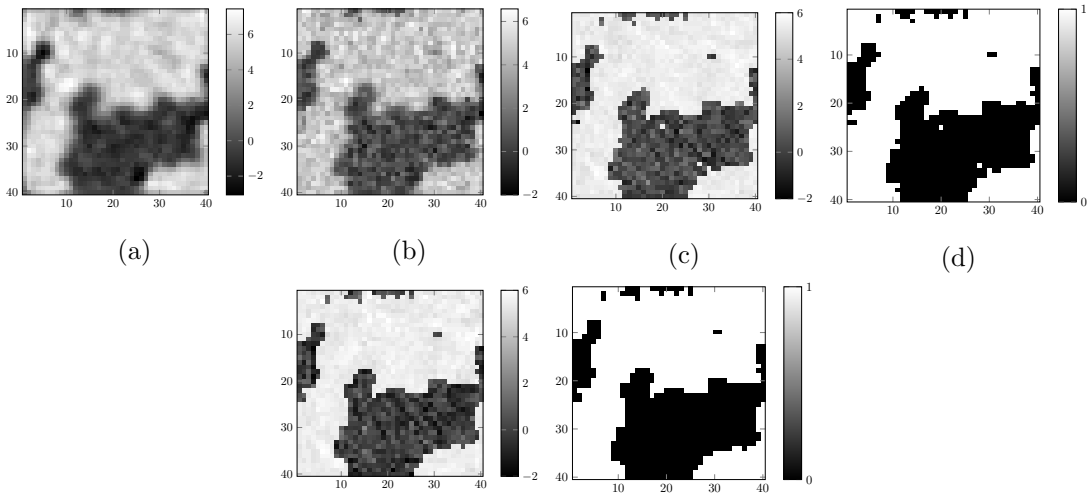


Figure 4.5: Estimated SLFs $\widehat{\mathbf{F}}$ at $\tau = 15$ (with 700 measurements) via (a) ridge-regularized LS ($\mu_f = 8.9 \times 10^{-4}$ and $\mathbf{C}_f = \mathbf{I}_{1,600}$); (b) TV-regularized LS ($\mu_f = 10^{-12}$); (c) Alg. 7 through (d) estimated hidden label field $\widehat{\mathbf{Z}}$; and (e) non-adaptive Bayesian algorithm, through (f) estimated $\widehat{\mathbf{Z}}$.

Table 4.2: True $\boldsymbol{\theta}$ and estimated $\hat{\boldsymbol{\theta}}$ via Alg. 7 (setting of Figs. 4.5c and 4.5d); and non-adaptive Bayesian algorithm (setting of Figs. 4.5e and 4.5f) averaged over 20 independent Monte Carlo runs.

$\boldsymbol{\theta}$	True	Est. (Alg. 7)	Est. (non-adaptive)
β	1.3	$1.309 \pm 2 \times 10^{-2}$	$1.309 \pm 3 \times 10^{-2}$
σ_ν^2	0.05	0.058 ± 10^{-2}	$0.053 \pm 1.3 \times 10^{-2}$
μ_{f_0}	0.2	$0.289 \pm 2 \times 10^{-2}$	$0.289 \pm 1.8 \times 10^{-2}$
μ_{f_1}	5	$4.996 \pm 7 \times 10^{-3}$	$4.996 \pm 7 \times 10^{-3}$
$\sigma_{f_0}^2$	1	$0.931 \pm 5 \times 10^{-2}$	$0.94 \pm 9.8 \times 10^{-2}$
$\sigma_{f_1}^2$	0.2	$0.198 \pm 2 \times 10^{-2}$	$0.193 \pm 2.8 \times 10^{-2}$

correspond to the leftmost points of the x-axis of Figs. 4.7a and 4.7b, respectively. The reconstruction performance is not severely degraded as σ_ν^2 increases, even in a high noise regime when $\sigma_\nu^2 = 10$, which suggests that the proposed algorithms are reasonably robust to measurement noise.

To assess the tracking capability of the proposed algorithm, slow variations in the SLF were simulated by introducing a moving object. The same setting used for Figs. 4.5c

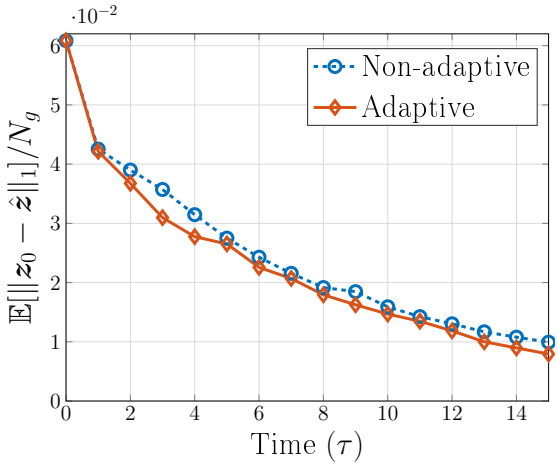


Figure 4.6: Progression of error in estimation of $\boldsymbol{z}$.

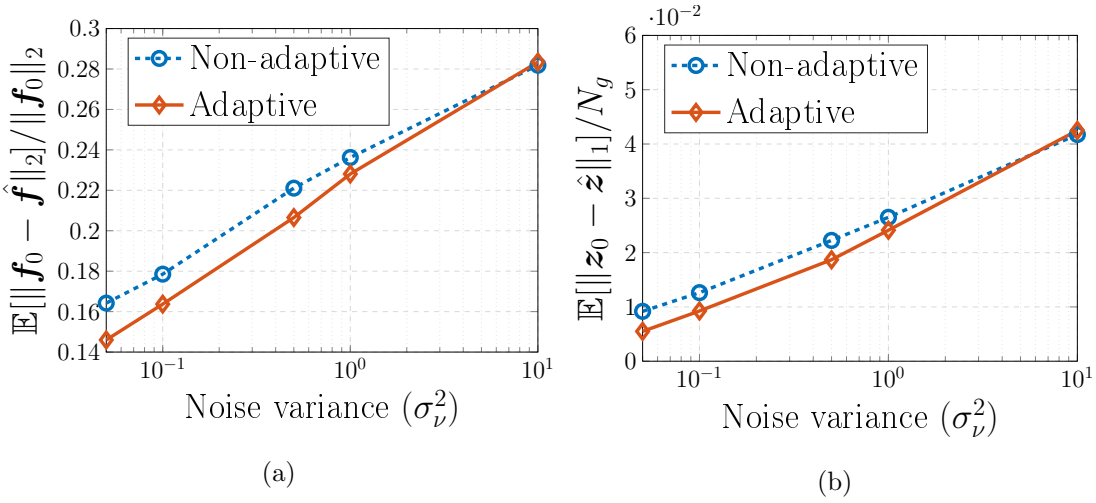
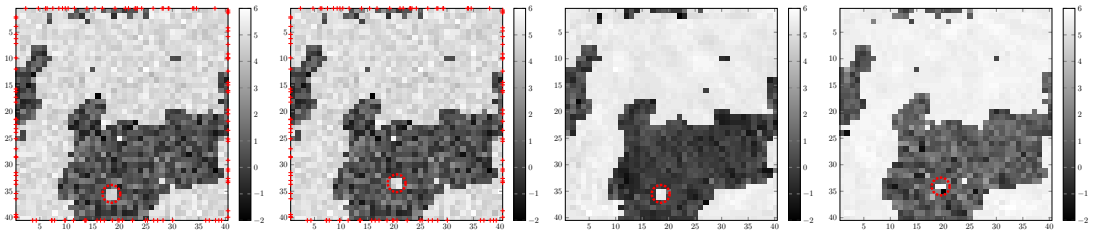


Figure 4.7: Reconstruction error vs. noise variance σ_ν^2 for (a) the SLF $\boldsymbol{f}$; and (b) the hidden label field $\boldsymbol{z}$.

and 4.5d was adopted. Measurements were generated with the SLF in Fig. 4.8a for $\tau = 0, \ldots, 5$, and that in Fig. 4.8b for the rest. The change in the SLF was assumed to happen once at $\tau = 6$. The reconstructed SLFs at $\tau = 5$ and $\tau = 15$ are shown in Figs. 4.8c and 4.8d, respectively. It is seen that only the SLF reconstructed at $\tau = 5$ correctly captures the moving object, while the stationary objects are estimated more clearly as τ increases, which reveals the trade-off between spatial and temporal



(a) (b) (c) (d)

Figure 4.8: True SLFs for (a) $\tau \in \{0, \ldots, 5\}$; and (b) $\tau \in \{6, \ldots, 15\}$; and estimated SLFs at (c) $\tau = 5$ (300 measurements); and (d) $\tau = 15$ (700 measurements) via Alg. 7. Dynamic objects are marked with dotted circles.

resolution.

The rest of this section tests the performance of the proposed algorithm in channel-gain estimation. To this end, the same setting used to produce Figs. 4.5c and 4.5d was adopted. From the estimate $\hat{\boldsymbol{f}}_{\text{MMSE}}$ obtained through Alg. 7, an estimate of the shadowing attenuation $\hat{s}(\mathbf{x}, \mathbf{x}')$ between two arbitrary points $\mathbf{x}$ and $\mathbf{x}'$ in $\mathcal{A}$ is obtained through (2.4) by replacing $\boldsymbol{f}$ with $\hat{\boldsymbol{f}}_{\text{MMSE}}$. Subsequently, an estimate of the channel-gain $\hat{g}(\mathbf{x}, \mathbf{x}')$ is obtained after substituting $\hat{s}(\mathbf{x}, \mathbf{x}')$ into (2.1).

Since g_0 and γ are known, obtaining $s(\mathbf{x}, \mathbf{x}')$ amounts to finding $g(\mathbf{x}, \mathbf{x}')$; cf. (2.1). This suggests adopting a performance metric quantifying the mismatch between $s(\mathbf{x}, \mathbf{x}')$ and $\hat{s}(\mathbf{x}, \mathbf{x}')$, using the normalized mean-square error

$$\text{NMSE} := \frac{\mathbb{E}\big[\int_{\mathcal{A}} \big(s(\mathbf{x}, \mathbf{x}') - \hat{s}(\mathbf{x}, \mathbf{x}')\big)^2 d\mathbf{x} d\mathbf{x}'\big]}{\mathbb{E}\big[\int_{\mathcal{A}} s^2(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}'\big]}$$

where the expectation is over the set $\{\mathbf{x}_n\}_{n=1}^N$ of sensor locations and realizations of $\{\nu_\tau\}_\tau$. Simulations estimated the expectations by averaging over 20 independent Monte Carlo runs. The integrals are approximated by averaging the integrand over 300 pairs of $(\mathbf{x}, \mathbf{x}')$ chosen independently and uniformly at random over the boundary of $\mathcal{A}$.

Fig. 4.9 compares the NMSE of the proposed method with those of the competing alternatives using the settings in Fig. 4.5. Evidently, the proposed method achieves the lowest NMSE for every τ. Observe that both Bayesian approaches outperform the regularized LS methods, which suggests the proposed method as a viable alternative of a conventional solution adopted for channel-gain cartography.

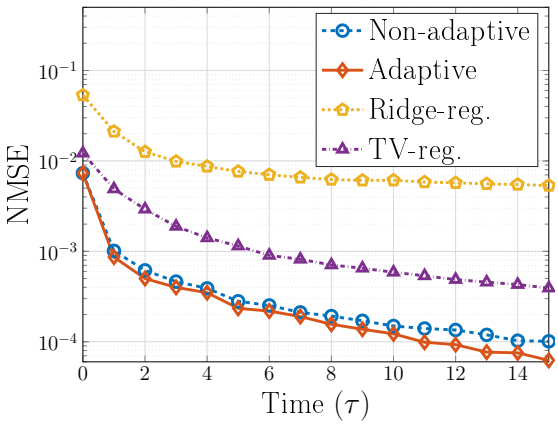


Figure 4.9: Progression of channel-gain estimation error.

Table 4.3: Hyper-hyperparameters of $\boldsymbol{\theta}$ for real data tests.

$\beta_{\max}$	m_0	m_1	$\sigma_k^2, \forall k$	a_ν	b_ν	$a_k, \forall k$	$b_k, \forall k$
2	0	1	0.01	1	0.01	0.01	0.01

4.3.2 Test with real data

This section validates the proposed method using the real data set in [33]. The test setup is depicted in Fig. 3.6, where $\mathcal{A} = [0.5, 20.5] \times [0.5, 20.5]$ is a square with sides of 20 feet (ft), over which a grid $\{\tilde{\mathbf{x}}_i\}_{i=1}^{1,681} := \{1, \ldots, 41\}^2$ of $N_g = 1,681$ points is defined. A collection of $N = 80$ sensors measure the channel attenuation at 2.425 GHz between pairs of sensor positions, marked with the $N = 80$ crosses in Fig. 3.6. To estimate g_0 and γ using the approach in [33], a first set of $2,400$ measurements was obtained before placing the artificial structure in Fig. 3.6. Estimates $\hat{g}_0 = 54.6$ (dB) and $\hat{\gamma} = 0.276$ were obtained during the calibration step. Afterwards, the structure comprising one pillar and six walls of different materials was assembled, and $T = 2,380$ measurements $\{\check{g}_{\tau'}\}_{\tau'=1}^{T}$ were acquired. Then, the calibrated measurements $\{\check{s}_{\tau'}\}_{\tau'=1}^{T}$ were obtained from $\{\check{g}_{\tau'}\}_{\tau'=1}^{T}$ by substituting $\hat{g}_0$ and $\hat{\gamma}$ into (2.5). In addition, $\{\mathbf{w}_{\tau'}\}_{\tau'=1}^{T}$ were constructed with w in (2.3) by using known locations of sensor pairs. Note that τ' is introduced to distinguish indices of the real data from τ used to index time slots in numerical tests.

We randomly selected $1,380$ measurements from $\{\check{s}_{\tau'}\}_{\tau'=1}^{T}$ to initialize $\check{\mathbf{s}}^{(0)}$, and used the remaining $1,000$ measurements to run the proposed algorithm under the

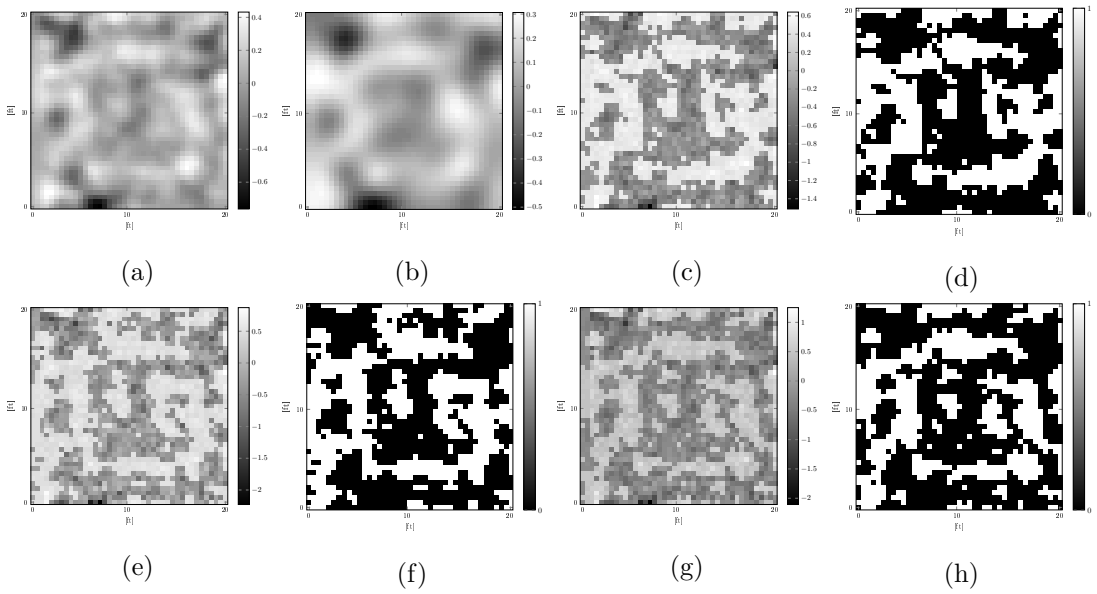


Figure 4.10: Estimated SLFs $\widehat{\mathbf{F}}$ at $\tau = 5$ (with $1,880$ measurements) via (a) ridge-regularized LS; (b) TV-regularized LS; (c) Alg. 7 through (d) estimated hidden label field $\widehat{\mathbf{Z}}$; and (e) non-adaptive Bayesian algorithm, through (f) estimated $\widehat{\mathbf{Z}}$, together with one-shot estimates (g) $\widehat{\mathbf{F}}_{\text{full}}$ and (h) $\widehat{\mathbf{Z}}_{\text{full}}$ obtained by using the full dataset (with $2,380$ measurements) via Alg. 7.

mini-batch operation for $\tau = 0, \ldots, 5$, where every $\mathcal{W}_{\tau+1}$ was formed by uniformly selecting $|\mathcal{W}_{\tau+1}| = 200$ weight vectors at random from $\{\mathbf{w}_{\tau'}\}_{\tau'}$ associated with the remaining $1,000$ measurements without replacement. Parameters of the proposed algorithm were set to, $N_{\text{CL}} = 2$, $N_{\text{Burn-in}} = 300$, $N_{\text{Iter}} = 1,000$, $\sigma_q^2 = 10^{-5}$, and the hyper-hyperparameters of $\boldsymbol{\theta}$ used are listed in Table 4.3. For initialization, $\boldsymbol{z}^{(0)}$ was found by drawing $z_i^{(0)} \sim \text{Bern}(0.5)$ $\forall i$. Vector $\boldsymbol{\theta}^{(0)}$ was set to have $\beta^{(0)} = 0.1$ and $\boldsymbol{\mu}_{f_k}^{(0)} = [m_0, m_1]^\top$, while σ_ν^2 and $\boldsymbol{\sigma}_{f_k}^2$ were initialized at random.

Following [2, 33], a spatial covariance matrix was used for $\mathbf{C}_f$ of the ridge-regularized LS estimator, which models the similarity between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ as $\big[\mathbf{C}_f\big]_{ij} = \sigma_s^2 \exp[-\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2/\kappa]$ [2], with $\sigma_s^2 = \kappa = 1$, and $\mu_f = 6 \times 10^{-2}$; see also [82]. On the other hand, the TV-regularized LS estimator was tested with $\mu_f = 4.3$ used in [82].

Fig. 4.10 displays estimated SLFs $\widehat{\mathbf{F}}$ and associated hidden fields $\widehat{\mathbf{Z}}$ at $\tau = 5$ obtained by the proposed method and its competing alternatives. The pattern of the artificial structure is clearly delineated on $\widehat{\mathbf{F}}$ in Fig. 4.10c estimated by the proposed method,

while the regularized LS estimators are not able to capture such pattern without post-processing of the estimated SLFs in Figs. 4.10a and 4.10b. Although the non-adaptive Bayesian algorithm reconstructed the visually satisfying SLF as shown in Fig. 4.10e, $\widehat{\mathbf{F}}$ from the proposed method depicts the artificial structure more clearly; see e.g., object patterns in Figs. 4.10c and 4.10e corresponding to the dry wall in Fig. 3.6. As a benchmark, an one-shot estimate of the SLF, denoted as $\widehat{\mathbf{F}}_{\text{full}}$, is also displayed in Fig. 4.10g, which was obtained via Alg. 7 by using the entire set of $2,380$ measurements. Comparison of $\widehat{\mathbf{F}}$ in Fig. 4.10c with $\widehat{\mathbf{F}}_{\text{full}}$ shows that the proposed algorithm enables one to reconstruct the SLF close to the benchmark by using fewer, but more informative measurements.

The second experiment investigated the efficacy of the proposed adaptive data acquisition method in estimating $\boldsymbol{z}$. By considering $\widehat{\mathbf{Z}}_{\text{full}} = \text{unvec}(\hat{\boldsymbol{z}}_{\text{full}})$ in Fig. 4.10h as a benchmark, the labeling error $\|\hat{\boldsymbol{z}}_{\text{full}} - \hat{\boldsymbol{z}}\|_1/N_g$ was used as a performance metric. Fig. 4.11 compares the labeling error of the proposed method with that of the non-adaptive algorithm, which are averaged over 20 independent Monte Carlo runs. The proposed method exhibits lower labeling errors than the non-adaptive one except when $\tau = 2$. This illustrates that the proposed data acquisition criterion delineates object patterns more accurately while also reducing the measurement collection cost.

To corroborate the hyperparameter estimation capability of the proposed algorithm, the estimates of $\boldsymbol{\theta}$ averaged over 20 independent Monte Carlo runs were listed in Table 4.4. The estimate $\hat{\boldsymbol{\theta}}$ obtained by using the full dataset was considered as a benchmark, to demonstrate that the proposed method estimates $\boldsymbol{\theta}$ closer to the benchmark. The scale of $\widehat{\sigma_\nu^2}$ in Table 4.4 is different from that in Table. 4.2. This can be explained by that the high noise level in $\{\check{s}_{\tau'}\}_{\tau'=1}^{T}$ due to the imperfect data calibration present in $\widehat{\sigma_\nu^2}$ to produce visually pleasing SLFs as shown in Fig. 4.10.

The last simulation assesses the performance of the proposed algorithm and competing alternatives for channel-gain cartography. The same set of shadowing measurements and simulation setup as in first simulations of this section were used. A channel-gain map is constructed to portray the gain between any point in the map, and a fixed receiver location $\mathbf{x}_{\text{rx}}$. Particularly, the proposed algorithm is executed and estimates $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\hat{\boldsymbol{f}}$ and w into (2.4). Subsequently, $\{\hat{g}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ into (2.1) with $\hat{g}_0$ and $\hat{\gamma}$.

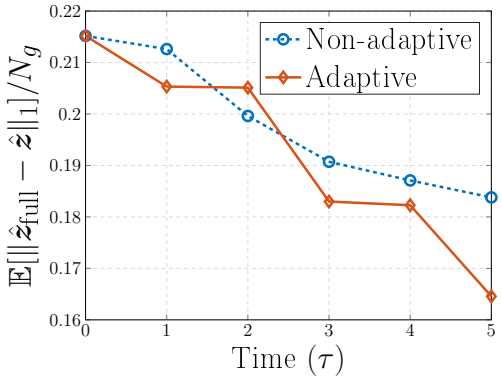


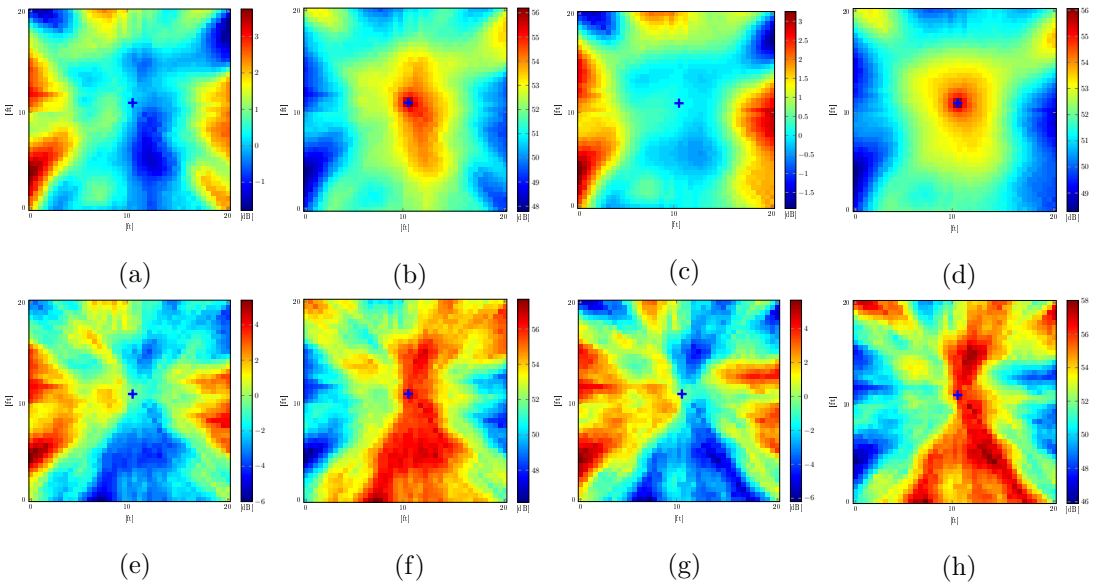
Figure 4.11: Progression of a mismatch between $\widehat{\boldsymbol{z}}$ and $\widehat{\boldsymbol{z}}_{\text{full}}$.

Table 4.4: Estimated $\hat{\boldsymbol{\theta}}$ via benchmark algorithm (setting of Figs. 4.10g and 4.10h); Alg. 7 (setting of Figs. 4.10c and 4.10d); and non-adaptive Bayesian algorithm (setting of Figs. 4.10e and 4.10f), averaged over 20 independent Monte Carlo runs.

$\boldsymbol{\theta}$	Est. (benchmark)	Est. (Alg. 7)	Est. (non-adaptive)
β	$0.499 \pm 2 \times 10^{-4}$	$0.5 \pm 5 \times 10^{-4}$	$0.5 \pm 6 \times 10^{-4}$
σ_ν^2	9.984 ± 0.05	10.60 ± 0.20	9.957 ± 0.23
μ_{f_0}	-0.275 ± 0.02	-0.278 ± 0.02	-0.301 ± 0.03
μ_{f_1}	0.463 ± 0.03	0.447 ± 0.03	0.504 ± 0.03
$\sigma_{f_0}^2$	0.629 ± 0.12	0.457 ± 0.13	0.456 ± 0.22
$\sigma_{f_1}^2$	0.171 ± 0.10	0.145 ± 0.10	0.325 ± 0.43

After defining $\hat{\boldsymbol{g}} := [\hat{g}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \ldots, \hat{g}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top$, one can construct the channel-gain map $\widehat{\mathbf{G}} := \text{unvec}(\hat{\boldsymbol{g}})$ with the receiver located at $\mathbf{x}_{\text{rx}}$.

Let $\widehat{\mathbf{S}} := \text{unvec}(\hat{\boldsymbol{s}})$ denote the shadowing map with $\hat{\boldsymbol{s}} := [\hat{s}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \ldots, \hat{s}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top$. Fig. 4.12 displays the estimated shadowing maps $\widehat{\mathbf{S}}$ and corresponding channel-gain maps $\widehat{\mathbf{G}}$, obtained via various methods, when the receiver is located at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked by the cross. In all channel-gain maps in Fig. 4.12, stronger attenuation is observed when a signal passes through either more building materials (bottom-right side of $\widehat{\mathbf{G}}$), or the concrete wall (left side of $\widehat{\mathbf{G}}$). In contrast, only the channel-gain maps in Figs. 4.12f and 4.12h reconstructed by the Bayesian methods exhibit less attenuation along the entrance of the artificial objects (top-right side of $\widehat{\mathbf{G}}$), while channel-gain tends to drop quickly within the vicinity of the receiver in the channel-gain maps obtained by the regularized LS estimators, as shown in Figs. 4.12b and 4.12d. This stems from the



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4.12: Estimated shadowing maps $\widehat{\mathbf{S}}$ and corresponding channel-gain maps $\widehat{\mathbf{G}}$ at $\tau = 5$ via (a)-(b) ridge-regularized LS (setting of Fig. 4.10a); (c)-(d) TV-regularized LS (setting of Fig. 4.10b); (e)-(f) Alg. 7 (setting of Fig. 4.10c); and (g)-(h) non-adaptive Bayesian algorithm (setting of Fig. 4.10e), with the receiver location at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked with the blue cross.

fact that free space and objects are more distinctively delineated in $\widehat{\mathbf{F}}$ by the Bayesian approaches. Note that slightly different observations were made in Figs. 4.12f and 4.12h since the shadowing map in Fig. 4.12g introduces stronger attenuation in free space below the receiver, which would disagree with intuition. All in all, the simulation results confirm that our approach could provide more specific CSI of the propagation medium, and thus endow the operation of cognitive radio networks with more accurate interference management.

4.4 Conclusion

This paper developed a novel channel-gain cartography algorithm that estimates the spatial loss field of the radio tomographic model, which is of interest in channel-gain

cartography and radio tomographic imaging applications, by using measurements adaptively collected based on the uncertainty sampling criterion. Different from conventional approaches, leveraging a hidden label field contributed to effectively account for inhomogeneities of the spatial loss field. The effectiveness of the novel algorithm was corroborated through extensive synthetic and real data experiments.

Chapter 5

A Variational Bayes Approach to Channel-gain Cartography

5.1 Motivation

To account for environmental heterogeneity, we introduced in Chapter 4 a Bayesian approach to learn a piecewise homogeneous spatial loss field (SLF) through a binary hidden Markov random field (MRF) model [38] via Markov chain Monte Carlo (MCMC) [30]. While the proposed MCMC solution outperformed competing alternatives in literature, this approach does not scale as the resolution of the SLF increases, because MCMC becomes computationally demanding due to resulting higher burn-in period for the sampler-convergence.

Aiming at efficient field estimators at affordable complexity, we will propose a variational Bayes (VB) framework for channel-gain cartography to approximate the analytically intractable MMSE or MAP estimators. Instead of considering the binary hidden MRF to model statistical heterogeneity in the SLF as in Chapter 4, we further generalize the SLF model by considering K-*ary* piecewise homogeneous regions for $K \geq 2$, to address a richer class of environmental heterogeneity. Besides developing efficient and affordable solutions for Bayesian channel-gain cartography, another contribution here is a data-adaptive sensor selection technique, with the goal of reducing uncertainty in the SLF, under the active learning framework [55]. Similar to Chapter 4, the conditional entropy of the SLF is considered as an uncertainty measure, giving rise to a novel sensor

selection criterion. Although this criterion is intractable especially when the size of the SLF and K are large, its efficient proxy can be obtained thanks to the availability of an approximate posterior model from the proposed VB algorithm.

5.2 Bayesian Model and Problem Formulation

Let $\mathcal{A}$ consist of K disjoint homogeneous regions $\mathcal{A}_k := \{\mathbf{x}|\mathbb{E}[f(\mathbf{x})] = \mu_{f_k}, \mathrm{Var}[f(\mathbf{x})] = \sigma^2_{f_k}\}$ for $k = 1, \ldots, K$, giving rise to a latent random label field $\boldsymbol{z} := [z(\tilde{\mathbf{x}}_1), \ldots, z(\tilde{\mathbf{x}}_{N_g})]^\top \in \{1, \ldots, K\}^{N_g}$ with K-ary entries $z(\tilde{\mathbf{x}}_i) = k$ if $\tilde{\mathbf{x}}_i \in \mathcal{A}_k \ \forall i, k$. The K separate regions will model heterogeneous environments. For such a paradigm, we model the conditional distribution of $f(\tilde{\mathbf{x}}_i)$ as

$$p(f(\tilde{\mathbf{x}}_i)|z(\tilde{\mathbf{x}}_i) = k) = \mathcal{N}(\mu_{f_k}, \sigma^2_{f_k}) \ \forall k \, . \tag{5.1}$$

We further assign the Potts prior to $\boldsymbol{z}$ in order to capture the dependency among spatially correlated labels. By the Hammersley-Clifford theorem [34], the Potts prior of $\boldsymbol{z}$ follows a Gibbs distribution

$$p(\boldsymbol{z}; \beta) = \frac{1}{C(\beta)} \exp\left[\sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i))\right] \tag{5.2}$$

where $\mathcal{N}(\tilde{\mathbf{x}}_i)$ is a set of indices comprising 1-hop neighbors of $\tilde{\mathbf{x}}_i$ on the rectangular grid in Fig. 4.1, β is a granularity coefficient controlling the degree of homogeneity in z, $\delta(\cdot)$ is Kronecker's delta, and the normalization constant

$$C(\beta) := \sum_{\boldsymbol{z} \in \mathcal{Z}} \exp\left[\sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i))\right] \tag{5.3}$$

is the partition function with $\mathcal{Z} := \{1, \ldots, K\}^{N_g}$. Note that the Ising prior in (4.3) in is a special case of (5.2) when $K = 2$. To ease exposition, β is assumed known or fixed a priori; see e.g., [19, 61, 50] for a means of estimating β. If $\{f(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g}$ are conditionally independent given $\boldsymbol{z}$, the model reduces to the Gauss-Markov-Potts model [5]. Such a model with $K = 3$ is depicted in Fig. 5.1 with the measurement model in (2.4).

Noise ν_t in (2.5) is assumed independent and identically distributed (i.i.d.) Gaussian with zero mean and variance σ_ν^2. Here, we correspondingly consider precisions of ν_t and $\{f_k\}_{k=1}^K$ that are denoted as $\varphi_\nu := 1/\sigma_\nu^2$ and $\varphi_{f_k} := 1/\sigma_{f_k}^2 \forall k$, respectively. Let also $\boldsymbol{\theta}$ be a hyperparameter vector comprising φ_ν and $\boldsymbol{\theta}_f := [\boldsymbol{\mu}_{f_k}^\top, \boldsymbol{\varphi}_f^\top]^\top$ with $\boldsymbol{\mu}_{f_k} := [\mu_{f_1}, \ldots, \mu_{f_K}]^\top \in \mathbb{R}^K$ and $\boldsymbol{\varphi}_f := [\varphi_{f_1}, \ldots, \varphi_{f_K}]^\top \in \mathbb{R}^K$. Assuming the independence among entries of $\boldsymbol{\theta}$, we deduce that

$$p(\boldsymbol{\theta}) = p(\varphi_\nu)p(\boldsymbol{\mu}_{f_k})p(\boldsymbol{\varphi}_f) = p(\varphi_\nu)\prod_{k=1}^K p(\mu_{f_k})p(\varphi_{f_k}) \tag{5.4}$$

where the priors $p(\varphi_\nu)$, $p(\boldsymbol{\mu}_{f_k})$, and $p(\boldsymbol{\varphi}_f)$ are as follows.

1) Noise precision φ_ν. With additive Gaussian noise having fixed mean, it is common to assign a conjugate prior to φ_ν that can reproduce a posterior in the same family of its prior. The gamma distribution for $\varphi_\nu \in \mathbb{R}^+$ serves this purpose, as

$$p(\varphi_\nu) = \mathcal{G}(a_\nu, b_\nu) := \frac{1}{\Gamma(a_\nu)b_\nu^{a_\nu}}(\varphi_\nu)^{a_\nu - 1}e^{-\varphi_\nu/b_\nu} \tag{5.5}$$

where a_ν is referred to as the shape parameter, b_ν as the scale parameter, and $\Gamma(\cdot)$ denotes the gamma function.

2) Hyperparameters $\boldsymbol{\theta}_f$ of the SLF. is the Gamma distribution parameterized by $\{a_k, b_k\}$; that is,

$$p(\mu_{f_k}) = \mathcal{N}(m_k, \sigma_k^2),\ k = 1, \ldots, K \tag{5.6}$$

$$p(\varphi_{f_k}) = \mathcal{G}(a_k, b_k), \quad k = 1, \ldots, K. \tag{5.7}$$

We stress that analytical tractability is the main motivation behind selecting the conjugate priors in (5.5)–(5.7).

Our goal of inferring $\boldsymbol{f}$, relies on the following posterior distribution that can be factored (within a constant) as

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{z}; \beta)p(\boldsymbol{\theta}) \tag{5.8}$$

where $p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu) \sim \mathcal{N}(\mathbf{W}_t^\top \boldsymbol{f}, \sigma_\nu^2 \mathbf{I}_t)$ is the data likelihood with the weight matrix $\mathbf{W}_t \in \mathbb{R}^{N_g \times t}$ formed from columns $\mathbf{w}_\tau^{(n,n')} := [w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_1), \ldots, w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_{N_g})]^\top :=$

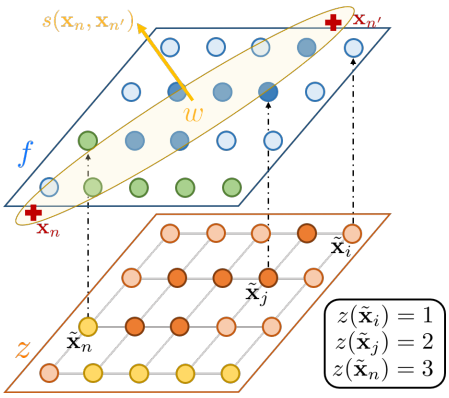


Figure 5.1: Gauss-Markov-Potts model for channel-gain cartography with $K = 3$, together with the measurement model for sensors located at $(\mathbf{x}_n, \mathbf{x}_{n'})$.

$[w_{\tau,1}, \ldots, w_{\tau,N_g}]^\top \in \mathbb{R}^{N_g}$ of the link $\mathbf{x}_{n(\tau)}$–$\mathbf{x}_{n'(\tau)}$ for $\tau = 1, \ldots, t$. Fig. 5.2 depicts our hierarchical Bayesian model for $\{\check{\mathbf{s}}_t, \boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}\}$ as a directed acyclic graph, where the dependence between (hyper) parameters is indicated with an arrow.

Given the posterior in (5.8), the conditional minimum mean-square error (MMSE) estimator of the field is

$$\hat{\boldsymbol{f}}_{\text{MMSE}} := \mathbb{E}[\boldsymbol{f}|\boldsymbol{z} = \hat{\boldsymbol{z}}_{\text{MAP}}, \check{\mathbf{s}}_t] \tag{5.9}$$

where the maximum a posteriori (MAP) label estimator is

$$\hat{\boldsymbol{z}}_{\text{MAP}} := \arg\max_{\boldsymbol{z}} p(\boldsymbol{z}|\check{\mathbf{s}}_t) \tag{5.10}$$

and the MMSE estimators of $\boldsymbol{\theta}$ entries are

$$\widehat{\varphi_\nu}_{\text{MMSE}} := \mathbb{E}[\varphi_\nu|\check{\mathbf{s}}_t] \tag{5.11}$$

$$\widehat{\mu_{f_k}}_{\text{MMSE}} := \mathbb{E}[\mu_{f_k}|\check{\mathbf{s}}_t], \;\; k = 1, \ldots, K \tag{5.12}$$

$$\widehat{\varphi_{f_k}}_{\text{MMSE}} := \mathbb{E}[\varphi_{f_k}|\check{\mathbf{s}}_t], \;\; k = 1, \ldots, K. \tag{5.13}$$

5.3 Channel-gain Cartography via variational Bayes

Although the estimator forms in (5.9)-(5.13) have been considered also in [42], obtaining estimates in practice is not tractable because the complex posterior in (4.10) is not

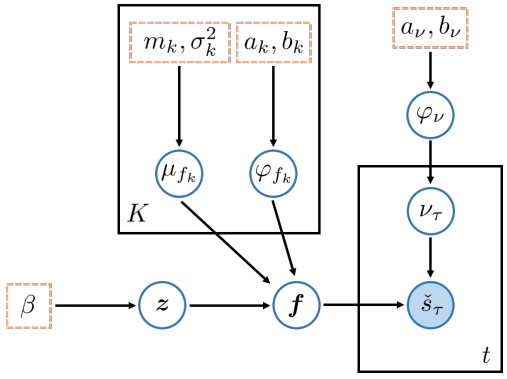


Figure 5.2: Graphical model representation of the hierarchical Bayesian model for (hyper) parameters (those in dashed boxes are fixed).

amenable to marginalization or maximization. To overcome this hurdle, one can resort to approximate Bayesian inference methods such as MCMC [30] that relies on samples of $\{\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}\}$ drawn from a complex distribution. Although MCMC can asymptotically approach an exact target distribution, such as the sought one in (4.10), it can be computationally demanding and thus does not scale well. Aiming at a scalable alternative, we will adopt the so-termed variational Bayes (VB) approach.

VB is a family of techniques to approximate a complex distribution by a tractable one termed variational distribution. A typical choice of an approximation criterion is to find the variational distribution q minimizing the Kullback-Leibler (KL) divergence ($D_{\mathrm{KL}}(q\|p)$) to a target distribution q. The variational distribution q is further assumed to belong to a certain family $\mathcal{Q}$ of distributions possessing a simpler form of dependence between variables than the original one; see also [70] for the so-termed mean-field approximation.

Tailored to the posterior in (4.10) the variational one, solves

$$\min_{q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\in\mathcal{Q}} D_{\mathrm{KL}}\left(q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\|p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_t)\right) \tag{5.14}$$

Using that $D_{\mathrm{KL}}\left(q\|p\right) := -\mathbb{E}_q[\ln(p/q)]$, the latter reduces to

$$\text{(P1)} \quad \max_{q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\in\mathcal{Q}} \underbrace{\mathbb{E}_{q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})}\left[\ln\left(\frac{p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\check{\mathbf{s}}_t)}{q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})}\right)\right]}_{=:\mathrm{ELBO}(q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}))} \tag{5.15}$$

where we drop the constant $p(\check{\mathbf{s}}_t)$ from the posterior that resulted in the so-termed *evidence lower bound* (ELBO) in (P1), which involves the joint $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t)$ factored as in the right-hand side (RHS) of (4.10). We choose the family $\mathcal{Q}$ as

$$\mathcal{Q} := \left\{ q : q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}) := q(\boldsymbol{f}|\boldsymbol{z}) q(\boldsymbol{z}) q(\boldsymbol{\theta}) = \prod_{i=1}^{N_g} q(f_i|z_i) \prod_{i=1}^{N_g} q(z_i) q(\boldsymbol{\theta}) \right\} \tag{5.16}$$

where $f_i := f(\tilde{\mathbf{x}}_i)$ and $z_i := z(\tilde{\mathbf{x}}_i)$ $\forall i$ for simplicity, and

$$q(\boldsymbol{\theta}) := q(\varphi_\nu) q(\boldsymbol{\mu}_{f_k}) q(\boldsymbol{\varphi}_f) = q(\varphi_\nu) \prod_{k=1}^{K} q(\mu_{f_k}) \prod_{k=1}^{K} q(\varphi_{f_k}). \tag{5.17}$$

Following the general VB steps [67], we will solve (P1) in (5.15) here via coordinate minimization among factors of $q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})$. Within a constant c, the optimal solutions have the form

$$\ln q^*(f_i|z_i) = \mathbb{E}_{-q(f_i|z_i)} \left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t) \right] + c \ \forall i \tag{5.18}$$

$$\ln q^*(z_i) = \mathbb{E}_{-q(z_i)} \left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t) \right] + c \ \forall i \tag{5.19}$$

$$\ln q^*(\boldsymbol{\theta}) = \mathbb{E}_{-q(\boldsymbol{\theta})} \left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t) \right] + c \tag{5.20}$$

where the expectation in (5.18) is over the variational pdf of $\boldsymbol{f}_{-i}$, $\boldsymbol{z}$, and $\boldsymbol{\theta}$, that is $\prod_{j \neq i} q(f_j|z_j) q(\boldsymbol{z}) q(\boldsymbol{\theta})$. Similar expressions are available for (5.19) and (5.20). The solutions in (5.18)–(5.20) are intertwined since the evaluation of one requires the others. We show in Appendices C.1–C.5 that the optimal solutions can be obtained iteratively; that is, per iteration $\ell = 1, 2, \ldots$, we have

$$q^{(\ell)}(f_i|z_i = k) = \mathcal{N}(\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i), \breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i)) \forall k \tag{5.21}$$

$$q^{(\ell)}(z_i = k) =: \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) = \frac{\breve{\breve{\zeta}}_k^{(\ell)}(\tilde{\mathbf{x}}_i)}{\sum_{k=1}^{K} \breve{\breve{\zeta}}_k^{(\ell)}(\tilde{\mathbf{x}}_i)} \forall k \tag{5.22}$$

$$q^{(\ell)}(\varphi_\nu) = \mathcal{G}(\breve{a}_\nu, \breve{b}_\nu^{(\ell)}) \tag{5.23}$$

$$q^{(\ell)}(\mu_{f_k}) = \mathcal{N}(\breve{m}_k^{(\ell)}, \breve{\sigma}_k^{2(\ell)}) \ \forall k \tag{5.24}$$

$$q^{(\ell)}(\varphi_{f_k}) = \mathcal{G}(\breve{a}_k^{(\ell)}, \breve{b}_k^{(\ell)}) \ \forall k \tag{5.25}$$

with variational parameters

$$\breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) = \left(\widetilde{\varphi}_{\nu}^{(\ell-1)} \sum_{\tau=1}^{t} w_{\tau,i}^2 + \widetilde{\varphi}_{f_k}^{(\ell-1)} \right)^{-1} \quad \forall k \tag{5.26}$$

$$\begin{aligned} \breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) &= \bar{f}_i^{(\ell-1)} \\ &\quad + \breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) \left[\left(\breve{m}_k^{(\ell-1)} - \bar{f}_i^{(\ell-1)} \right) \widetilde{\varphi}_{f_k}^{(\ell-1)} + \widetilde{\varphi}_{\nu}^{(\ell-1)} \sum_{\tau=1}^{t} w_{\tau,i} \left(\check{s}_\tau - \overline{s}_\tau^{(\ell-1)} \right) \right] \forall k \end{aligned} \tag{5.27}$$

$$\begin{aligned} \breve{\breve{\zeta}}_k^{(\ell)}(\tilde{\mathbf{x}}_i) &= \exp \Bigg\{ - \frac{\widetilde{\varphi}_{f_k}^{(\ell-1)}}{2} \Bigg[\breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 - 2\breve{m}_k^{(\ell-1)} \breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) + \breve{\sigma}_k^{2(\ell-1)} \\ &\quad + \left(\breve{m}_k^{(\ell-1)} \right)^2 \Bigg] + \frac{1}{2} \left[\psi \left(\breve{a}_k^{(\ell-1)} \right) + \ln \breve{b}_k^{(\ell-1)} \right] + \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \breve{\zeta}_k^{(\ell-1)}(\tilde{\mathbf{x}}_j) \Bigg\} \forall k \end{aligned} \tag{5.28}$$

$$\breve{a}_\nu = a_\nu + \frac{t}{2} \tag{5.29}$$

$$\begin{aligned} \breve{b}_\nu^{(\ell)} &= \Bigg\{ \frac{1}{b_\nu} + \frac{1}{2} \sum_{\tau=1}^{t} \check{s}_\tau^2 - 2\check{s}_\tau \overline{s}_\tau^{(\ell)} + \sum_{i=1}^{N_g} w_{\tau,i}^2 \Bigg[\sum_{k=1}^{K} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \left(\breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 \right) \\ &\quad - \left(\bar{f}_i^{(\ell)} \right)^2 \Bigg] + \left(\overline{s}_\tau^{(\ell)} \right)^2 \Bigg\}^{-1} \end{aligned} \tag{5.30}$$

$$\breve{\sigma}_k^{2(\ell)} = \left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \widetilde{\varphi}_{f_k}^{(\ell-1)} \right)^{-1} \quad \forall k \tag{5.31}$$

$$\breve{m}_k^{(\ell)} = \breve{\sigma}_k^{2(\ell)} \left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \widetilde{\varphi}_{f_k}^{(\ell-1)} \breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right) \forall k \tag{5.32}$$

$$\breve{a}_k^{(\ell)} = a_k + \frac{1}{2} \sum_{i=1}^{N_g} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \forall k \tag{5.33}$$

$$\begin{aligned} \breve{b}_k^{(\ell)} &= \Bigg[\frac{1}{b_k} + \frac{1}{2} \sum_{i=1}^{N_g} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \Big(\breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 - 2\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \breve{m}_k^{(\ell-1)} + \breve{\sigma}_k^{2(\ell-1)} \\ &\quad + \left(\breve{m}_k^{(\ell-1)} \right)^2 \Big) \Bigg]^{-1} \forall k \end{aligned} \tag{5.34}$$

where $\psi(\cdot)$ is the digamma function, $\bar{f}_i^{(\ell)} := \sum_{k=1}^{K} \breve{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \; \forall i$, $\overline{s}_\tau^{(\ell)} := \sum_{i=1}^{N_g} w_{\tau,i} \bar{f}_i^{(\ell)}$

$\forall \tau$, $\widetilde{\varphi}_{\nu}^{(\ell)} := \mathbb{E}_{q^{(\ell)}(\varphi_\nu)}[\varphi_\nu] = \breve{a}_\nu \breve{b}_\nu^{(\ell)}$, and $\widetilde{\varphi}_{f_k}^{(\ell)} := \mathbb{E}_{q^{(\ell)}(\varphi_{f_k})}[\varphi_{f_k}] = \breve{a}_k^{(\ell)} \breve{b}_k^{(\ell)}$ $\forall k$; see Appendices C.1–C.5 for detailed derivation of the variational factors and parameters in (5.21)–(5.34).

Upon convergence of the iterative solvers, the (approximate) MAP estimator of $\boldsymbol{z}$ can be obtained as

$$\hat{z}_{\mathrm{MAP},i} = \arg \max_{z_i \in 1,\ldots,K} q^*(z_i) \ \forall i, \tag{5.35}$$

and then the (approximate) MMSE estimator of $\boldsymbol{f}$ as

$$\hat{f}_{\mathrm{MMSE},i} \simeq \mathbb{E}_{q^*(f_i|\hat{z}_{\mathrm{MAP},i})}[f_i] = \breve{\mu}^*_{f_{\hat{z}_{\mathrm{MAP},i}}}(\tilde{\mathbf{x}}_i) \ \forall i \tag{5.36}$$

while $\boldsymbol{\theta}$ is estimated using the marginal MMSE estimators

$$\widehat{\varphi_\nu}_{\mathrm{MMSE}} \simeq \mathbb{E}_{q^*(\varphi_\nu)}[\varphi_\nu] = \breve{a}^*_\nu \breve{b}^*_\nu \tag{5.37}$$

$$\widehat{\mu_{f_k}}_{\mathrm{MMSE}} \simeq \mathbb{E}_{q^*(\mu_{f_k})}[\mu_{f_k}] = \breve{m}^*_k \ \forall k \tag{5.38}$$

$$\widehat{\varphi_{f_k}}_{\mathrm{MMSE}} \simeq \mathbb{E}_{q^*(\varphi_{f_k})}[\varphi_{f_k}] = \breve{a}^*_k \breve{b}^*_k \ \forall k. \tag{5.39}$$

The VB algorithm to obtain $\{\hat{f}_{\mathrm{MMSE},i}\}_{i=1}^{N_g}$, $\{\hat{z}_{\mathrm{MAP},i}\}_{i=1}^{N_g}$, $\hat{\boldsymbol{\theta}}_{\mathrm{MMSE}}$, and $q^*(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})$ is tabulated in Alg. 8.

Remark 5.1 (Assessing convergence). The steps of Alg. 8 guarantee that the ELBO monotonically increases across iterations ℓ [7]. Hence, convergence of the solution can be assessed by monitoring the change in the ELBO of (P1) in (5.15), which for a preselected threshold $\xi > 0$ suggests stopping at iteration ℓ if $\mathrm{ELBO}\left(q^{(\ell)}(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})\right) - \mathrm{ELBO}\left(q^{(\ell-1)}(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})\right) \leq \xi$.

Remark 5.2 (Computational complexity). For Alg. 8, the complexity order to update $q(f_i|z_i = k)$ $\forall i, k$ per iteration ℓ is $\mathcal{O}(tKN_g)$ to compute $\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)$ in (5.27), while updating $\breve{\zeta}_k(\tilde{\mathbf{x}}_i)$ $\forall i, k$ via (5.22) incurs complexity $\mathcal{O}(KN_g)$. In addition, updating $q(\boldsymbol{\theta})$ has complexity $\mathcal{O}(tKN_g)$ that is dominated by the computation of $\breve{b}_\nu$ in (5.30). Overall, the per-iteration complexity of Alg. 8 is $\mathcal{O}\left((2t+1)KN_g\right)$.

Note that a sample-based counterpart of Alg. 8 via MCMC in [50] incurs complexity in the order of $\mathcal{O}(N_g^3)$. For conventional methods to estimate $\boldsymbol{f}$, the ridge regularized LS [33] has a one-shot (non-iterative) complexity of $\mathcal{O}(N_g^3)$, while the total variation (TV) regularized LS via the alternating direction method of multipliers (ADMM) in [73]

Algorithm 8 Field estimation via variational Bayes

Input: $\check{\mathbf{s}}_t$, $\mathbf{W}_t$, $\left\{a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}$, and N_{Iter}.

Initialize $q^{(0)}(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})$ and set $\ell = 0$
Obtain $\breve{a}_\nu$ with (5.29)
while ELBO has not converged and $\ell \leq N_{\text{Iter}}$ **do**
  Set $\ell \leftarrow \ell + 1$
  Obtain $\breve{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i)\ \forall i, k$ via (5.26)
  Obtain $\breve{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i)\ \forall i, k$ via (5.27)
  Obtain $q^{(\ell)}(z_i = k)\ \forall i, k$ via (5.22)
  Obtain $\breve{b}_\nu^{(\ell)}$ via (5.30)
  Obtain $\breve{\sigma}_k^{2(\ell)}\ \forall k$ via (5.31)
  Obtain $\breve{m}_k^{(\ell)}\ \forall k$ via (5.32)
  Obtain $\breve{a}_k^{(\ell)}\ \forall k$ via (5.33)
  Obtain $\breve{b}_k^{(\ell)}\ \forall k$ via (5.34)
end while
Set $q^*(f_i|z_i) = q^{(\ell)}(f_i|z_i)$ and $q^*(z_i) = q^{(\ell)}(z_i)\ \forall i$
Set $q^*(\boldsymbol{\theta}) = q^{(\ell)}(\boldsymbol{\theta})$
Estimate $\hat{z}_{\text{MAP},i} = \arg\max_{z_i \in \{1,\ldots,K\}} q^*(z_i)\ \forall\ i$
Estimate $\hat{f}_{i,\text{MMSE}} = \breve{\mu}^*_{f_{\hat{z}_{\text{MAP},i}}}(\tilde{\mathbf{x}}_i)\ \forall\ i$
Estimate $\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbb{E}_{q^*(\boldsymbol{\theta})}[\boldsymbol{\theta}]$ via (5.37)–(5.39)
return $\hat{\boldsymbol{f}}_{\text{MMSE}}$, $\hat{\boldsymbol{z}}_{\text{MAP}}$, $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$, $q^*(\boldsymbol{f}|\boldsymbol{z})$, $q^*(\boldsymbol{z})$, and $q^*(\boldsymbol{\theta})$

incurs complexity of $\mathcal{O}(N_g^3)$ per iteration ℓ; see also [51, 82] for details. This means that Alg. 8 incurs the lowest per-iteration complexity, which becomes more critical as N_g increases.

5.4 Data-adaptive Sensor Selection

Here we deal with cost-effective channel-gain cartography as new data are collected by interactively querying the location of sensing radios to acquire a minimal but most informative measurements. To this end, a measurement (or a mini-batch of measurements) can be adaptively collected using a set of available sensing radio pairs, with the goal of reducing the uncertainty of $\boldsymbol{f}$. Since the proposed Bayesian framework accounts for the uncertainty through $\breve{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i)$ in (5.32), we adopt the conditional entropy [17] to serve as

an uncertainty measure of $\boldsymbol{f}$ at time slot τ, namely,

$$H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)=\sum_{\boldsymbol{z}'\in\mathcal{Z}}\int p(\boldsymbol{z}',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)d\check{\mathbf{s}}'_\tau, \tag{5.40}$$

where $\hat{\boldsymbol{\theta}}_\tau$ is the estimate obtained via (5.37)–(5.39) per slot τ, and

$$\begin{aligned}H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)&:=-\int p(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)\ln p(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)d\boldsymbol{f}\\&=\frac{1}{2}\ln\left|\boldsymbol{\Sigma}_{f|z',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau}\right|+\frac{N_g}{2}\left(1+\ln 2\pi\right)\end{aligned} \tag{5.41}$$

as $p(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)$ is Gaussian with covariance matrix $\boldsymbol{\Sigma}_{f|z,\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau}:=\left(\widehat{\varphi_\nu}\mathbf{W}_\tau\mathbf{W}_\tau^\top+\widehat{\boldsymbol{\Phi}_{f|z}}\right)^{-1}$ with $\widehat{\boldsymbol{\Phi}_{f|z}}:=\mathrm{diag}\left(\{\widehat{\varphi_{f_{z_i}}}\}_{i=1}^{N_g}\right)$ [50]. Then, using the matrix determinant identity lemma [35, Chap. 18], it is not hard to show that

$$\begin{aligned}H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_{\tau+1};\hat{\boldsymbol{\theta}}_\tau)=H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)-\frac{1}{2}\sum_{\boldsymbol{z}'\in\mathcal{Z}}\int p(\boldsymbol{z}',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)\\\times\ln\left(1+\widehat{\varphi_\nu}{\mathbf{w}_{\tau+1}^{(n,n')}}^\top\boldsymbol{\Sigma}_{f|z',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau}\mathbf{w}_{\tau+1}^{(n,n')}\right)d\check{\mathbf{s}}'_\tau.\end{aligned} \tag{5.42}$$

To obtain $\check{s}_{\tau+1}$, we choose a pair of sensors (n^*,n'^*), or equivalently find $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ minimizing $H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_{\tau+1};\hat{\boldsymbol{\theta}}_\tau)$.

Given $\check{\mathbf{s}}_\tau$, we then find $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ by solving

$$\text{(P2)}\qquad\max_{\substack{\mathbf{w}_{\tau+1}^{(n,n')}:\\(n,n')\in\mathcal{M}_{\tau+1}}}\ \mathbb{E}_{p(\boldsymbol{z}|\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)}\left[h(\boldsymbol{z},\check{\mathbf{s}}_\tau,\mathbf{w}_{\tau+1}^{(n,n')};\hat{\boldsymbol{\theta}}_\tau)\right] \tag{5.43}$$

where $h(\boldsymbol{z},\check{\mathbf{s}}_\tau,\mathbf{w};\hat{\boldsymbol{\theta}}_\tau):=\ln\left(1+\widehat{\varphi_\nu}\mathbf{w}^\top\boldsymbol{\Sigma}_{f|z,\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau}\mathbf{w}\right)$ and $\mathcal{M}_\tau:=\left\{(n,n')|\exists(\mathbf{x}_n\text{–}\mathbf{x}_{n'})\text{ at }\tau,(n,n')\in\{1,\ldots,N\}^2\right\}$ denotes the set of available sensing radio pairs at slot τ.

Clearly, (P2) in (5.43) cannot be directly solved because $p(\boldsymbol{z}|\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)$ is not tractable e.g., by marginalizing the posterior in (5.8). Hence, evaluating the cost of (P2) is intractable for large N_g as $|\mathcal{Z}|=2^{N_g}$. Fortunately, we show next how (P2) can be approximately reformulated using the variational distribution $q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})$. Consider first

that

$$p(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau,\boldsymbol{\theta}) = \frac{p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_\tau)}{p(\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_\tau)} \approx \frac{q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})}{q(\boldsymbol{z},\boldsymbol{\theta})} = q(\boldsymbol{f}|\boldsymbol{z}), \tag{5.44}$$

which yields the approximation of $H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)$ in (5.41), as

$$H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}',\check{\mathbf{s}}_\tau=\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau) \approx \frac{1}{2}\ln\left|\breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau}\right| + \frac{N_g}{2}\left(1+\ln 2\pi\right) \tag{5.45}$$

with $\breve{\boldsymbol{\Sigma}}_{f|z,\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau} := \text{diag}\left(\{\breve{\sigma}^2_{f_{z_i}}(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g}\right)$; and subsequently, that of $H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)$ by substituting (5.45) into (5.40).

Similar to (5.42), we then show in Appendix C.6 that

$$\begin{aligned} H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_{\tau+1};\hat{\boldsymbol{\theta}}_\tau) \approx\ & H(\boldsymbol{f}|\boldsymbol{z},\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau) \\ & - \frac{1}{2}\sum_{\boldsymbol{z}'\in\mathcal{Z}}\int p(\boldsymbol{z}',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau)\ln\left|\mathbf{I}_{N_g}+\widetilde{\varphi}_\nu\boldsymbol{\Delta}_{w_{\tau+1}}\breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_\tau;\hat{\boldsymbol{\theta}}_\tau}\right| d\check{\mathbf{s}}'_\tau \end{aligned} \tag{5.46}$$

where $\boldsymbol{\Delta}_{w_{\tau+1}} := \text{diag}\left(\mathbf{w}_{\tau+1}^{(n,n')}\circ\mathbf{w}_{\tau+1}^{(n,n')}\right)$, with $\circ$ denoting the Hadamard product. Given $\check{\mathbf{s}}_\tau$, and using the approximation $p(\boldsymbol{z}|\check{\mathbf{s}}_\tau;\hat{\boldsymbol{\theta}}_\tau)\approx q(\boldsymbol{z})$, we can reformulate (P2) as (cf. (5.46))

$$\text{(P2')}\qquad \max_{\substack{\mathbf{w}_{\tau+1}^{(n,n')}:\\ (n,n')\in\mathcal{M}_{\tau+1}}} \underbrace{\sum_{i=1}^{N_g}\mathbb{E}_{q(z_i)}\left[\ln\left(1+\widetilde{\varphi}_\nu\breve{\sigma}^2_{f_{z_i}}(\tilde{\mathbf{x}}_i)w^2_{\tau+1,i}\right)\right]}_{=:\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})}.$$

Solving (P2') using a greedy search, we obtain the pair of sensors (n^*,n'^*) associated with $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$, based on which we collect the informative measurement $\check{s}_{\tau+1}$.

The overall algorithm for adaptive channel-gain cartography via VB is tabulated in Alg. 9.

Remark 5.3 (Mini-batch setup). The proposed data-adaptive sensor selection scheme can be easily extended to a mini-batch setup of size N_{Batch} per time slot τ as follows: i) find weight vectors $\left\{\mathbf{w}_{\tau+1}^{(n^{(m)},n'^{(m)})}\right\}_{m=1}^{N_{\text{Batch}}}$ for $\left\{\left(n^{(m)},n'^{(m)}\right)\right\}_{m=1}^{N_{\text{Batch}}}\subset\mathcal{M}_{\tau+1}$ associated with N_{Batch} largest values of $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (P2'), and collect $\{\check{s}_{\tau+1}^{(m)}\}_{m=1}^{N_{\text{Batch}}}$ from pairs of sensors revealed from those weight vectors (steps 4–5 in Alg. 9); and ii) aggregate those measurements below $\check{\mathbf{s}}_\tau$ to construct $\check{\mathbf{s}}_{\tau+1} := [\check{\mathbf{s}}_\tau^\top,\check{s}_{\tau+1}^{(1)},\ldots,\check{s}_{\tau+1}^{(N_{\text{Batch}})}]^\top$ (step 6 in Alg. 9). Numerical tests are presented next to assess the mini-batch operation

Algorithm 9 Adaptive channel-gain cartography via variational Bayes

Input: $\check{\mathbf{s}}^{(0)}$, $\mathbf{W}^{(0)}$, $\left\{a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}$, and N_{Iter}.

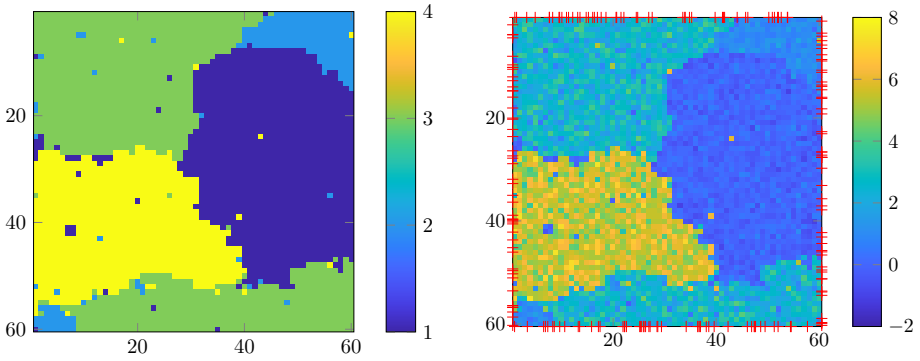
1: Set $\check{\mathbf{s}}_0 = \check{\mathbf{s}}^{(0)}$ and $\mathbf{W}_0 = \mathbf{W}^{(0)}$
2: **for** $\tau = 0, 1, \ldots$ **do**
3: Obtain $\hat{\boldsymbol{f}}_{\text{MMSE}}$, $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$, and $q^*(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})\}$ via Alg. 8$\left(\check{\mathbf{s}}_\tau, \mathbf{W}_\tau, \left\{a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}, N_{\text{Iter}}\right)$
4: Evaluate $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (P2') $\forall \{n, n'\} \in \mathcal{M}_{\tau+1}$
5: Collect $\check{s}_{\tau+1}$ from (n^*, n'^*) with $\max \bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$
6: Set $\check{\mathbf{s}}_{\tau+1} = [\check{\mathbf{s}}_\tau^\top, \check{s}_{\tau+1}]^\top$ and $\mathbf{W}_{\tau+1} = [\mathbf{W}_\tau, \mathbf{w}_{\tau+1}^{(n^*,n'^*)}]$
7: **end for**
8: Specify arbitrary locations of interest $\{\mathbf{x}, \mathbf{x}'\} \in \mathcal{A}$
9: Estimate $\hat{s}(\mathbf{x}, \mathbf{x}')$ via (2.4) by using $\hat{\boldsymbol{f}}_{\text{MMSE}}$
10: Estimate $\hat{g}(\mathbf{x}, \mathbf{x}')$ via (2.1) by using g_0, γ, and $\hat{s}(\mathbf{x}, \mathbf{x}')$

of Alg. 9.

5.5 Numerical Tests

Performance of the proposed algorithms was assessed through numerical tests using MATLAB on synthetic and real datasets. Comparisons were carried out with existing methods, including the ridge-regularized SLF estimate given by $\hat{\boldsymbol{f}}_{\text{LS}} = (\mathbf{W}_t\mathbf{W}_t^\top + \mu_f\mathbf{C}_f^{-1})^{-1}\mathbf{W}_t\check{\mathbf{s}}_t$ [33], where $\mathbf{C}_f$ is a spatial covariance matrix modeling the similarity between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ in area $\mathcal{A}$. We further tested the TV-regularized LS scheme in [73], which solves the problem in (4.1) with $\mathcal{R}(\boldsymbol{f})$ in (4.48). We also tested an MCMC-based counterpart of Alg. 9 for estimating the posterior in (4.10), and solving (P2) in (5.43); see e.g., [50, 72] for details.

We further compared the proposed data-adaptive sensor selection with simple *random sampling* for both regularized LS estimators, by selecting $\left\{\left(n^{(m)}, n'^{(m)}\right)\right\}_{m=1}^{N_{\text{Batch}}}$ uniformly at random to collect $\{\check{s}_{\tau+1}^{(m)}\}_{m=1}^{N_{\text{Batch}}}$ $\forall\tau$. Alg. 9 after replacing steps 4–5 with random sampling is termed *non-adaptive VB algorithm*, and will be compared with the proposed method throughout synthetic and real data tests.



(a) (b)

Figure 5.3: True fields for synthetic tests: (a) hidden label field $\mathbf{Z}_0$ and (b) spatial loss field $\mathbf{F}_0$ with $N = 200$ sensor locations marked with crosses.

5.5.1 Test with synthetic data

This section validates the proposed algorithm using synthetic datasets. Random tomographic measurements were collected from $N = 200$ sensors uniformly deployed on the boundary of $\mathcal{A} := [0.5, 60.5] \times [0.5, 60.5]$. Using these measurements, the SLF was reconstructed over the grid $\{\mathbf{x}_i\}_{i=1}^{3,600} := \{1, \ldots, 60\}^2$. To generate the ground-truth SLF $\boldsymbol{f}_0$, the ground-truth label field $\boldsymbol{z}_0$ was generated via Gibbs sampling [29] by using the Potts prior of $\boldsymbol{z}$ in (5.2) with $\beta = 1.5$ and $K = 4$. Given $\boldsymbol{\theta}_f := [\boldsymbol{\mu}_{f_k}^\top, \boldsymbol{\varphi}_f^\top]^\top$ with $\boldsymbol{\mu}_{f_k} = [0, 1, 2.5, 5.5]^\top$ and $\boldsymbol{\varphi}_f = [10, 10, 2, 2]^\top$, vector $\boldsymbol{f}_0$ was constructed to have $f(\mathbf{x}_i) \sim \mathcal{N}(\mu_{f_k}, \varphi_{f_k}^{-1})$ $\forall \mathbf{x}_i \in \mathcal{A}_k$, $\forall k$ conditioned on the labels in $\boldsymbol{z}_0$. The resulting hidden label field $\mathbf{Z}_0 := \text{unvec}(\boldsymbol{z}_0) \in \{1, 2, 3, 4\}^{60\times 60}$, and the true SLF $\mathbf{F}_0 := \text{unvec}(\boldsymbol{f}_0) \in \mathbb{R}^{60\times 60}$ are depicted in Fig. 5.3 with sensor locations marked by crosses. The effects of calibration are not accounted for, meaning that g_0 and γ are assumed to be known, and the fusion center directly uses shadowing measurements $\check{\mathbf{s}}_\tau$. Under the mini-batch operation, each measurement $\check{s}_\tau^{(m)}$ $\forall \tau, m$ was generated according to (2.5), where s_τ was obtained using (2.4) with w set to the normalized ellipse model in (2.3) with $\lambda = 0.39$, while ν_τ was set to follow a zero-mean Gaussian pdf with $\varphi_\nu = 20$. To construct $\mathcal{M}_{\tau+1}$ per time slot τ, $|\mathcal{M}_{\tau+1}| = 200$ pairs of sensors were uniformly selected at random with replacement. Then, $N_{\text{Batch}} = 100$ shadowing measurements were collected at $\left\{ \left(n^{(m)}, n'^{(m)}\right) \right\}_{m=1}^{N_{\text{Batch}}} \subset \mathcal{M}_{\tau+1}$ to run Alg. 9 for $\tau = 0, 1, \ldots, 8$.

In all synthetic tests, the simulation parameters were set to $N_{\text{Iter}} = 3,000$ and $\xi = 10^{-6}$; hyper-hyper parameters of ν_t were set to $a_\nu = 1,300$ and $b_\nu = 2$; and those of $\boldsymbol{\theta}_f$ were set as listed in Table. 5.1. To execute Alg. 8, variational parameters of $q^{(0)}(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})$ were initialized as follows: $\left\{\breve{\mu}_{f_k}^{(0)}(\tilde{\mathbf{x}}_i)\right\}_{i=1}^{N_g}$ $\forall k$, $\breve{b}_\nu^{(0)}$, $\left\{\breve{\sigma}_k^{2(0)}\right\}_{k=1}^4$, and $\left\{\breve{a}_k^{(0)}, \breve{b}_k^{(0)}\right\}_{k=1}^4$ were drawn from the uniform distribution $\mathcal{U}(0,1)$, while $\breve{m}_k^{(0)} = m_k$ $\forall k$; and it was set to $\breve{\zeta}_k^{(0)}(\tilde{\mathbf{x}}_i) = 1/4$ $\forall i, k$. Furthermore, $\check{\mathbf{s}}^{(0)}$ was collected from 800 pairs of sensors selected at random, which determined $\mathbf{W}^{(0)}$. To find μ_f of the competing alternatives, the L-curve [49, Chapter 26] was used for the ridge regularization, while the generalized cross-validation [31] was adopted for the TV regularization. The hyper-hyper parameters of $\boldsymbol{\theta}$ used for the proposed algorithm were also adopted to run its MCMC-based counterpart.

The first experiment is performed to validate Alg. 9. Estimates of SLFs $\widehat{\mathbf{F}} := \text{unvec}(\hat{\boldsymbol{f}})$ and the associated hidden label fields $\widehat{\mathbf{Z}} := \text{unvec}(\hat{\boldsymbol{z}})$ at time slot $\tau = 8$ obtained via Alg. 9, and the competing alternatives, are depicted in Figs. 5.4a–5.4j. One-shot estimates of the SLF and associated hidden field, denoted as $\widehat{\mathbf{F}}_{\text{full}}$ and $\widehat{\mathbf{Z}}_{\text{full}}$, respectively, are also displayed in Figs. 5.4k and 5.4l, which were obtained via Alg. 9 by using the entire set of $2,400$ measurements collected till $\tau = 8$. Clearly, satisfactory results were obtained only by teh approximate Bayesian inference methods including MCMC and VB because every piecewise homogeneous region was accurately classified through the hidden label field. As discussed in Remark 5.2 however, the proposed algorithm is computationally much more efficient than the ones using MCMC. Per-iteration execution time was 0.04 (sec) for Alg. 9 on average, while that was 3.64 (sec) for the MCMC method. On the other hand, the regularized LS solutions were unable to accurately reconstruct the SLF, as depicted in Figs. 5.4a and 5.4b.

To test the proposed sensor selection method, $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{Z}}$ found using the non-adaptive VB algorithm are depicted in Figs. 5.4e and 5.4f. Visual comparison of Figs. 5.4c and 5.4e reveals that the reconstruction performance for $\mathbf{F}$ can be improved with the same number of measurements by adaptively selecting pairs of sensors. Accuracy of $\hat{\boldsymbol{z}}$ was also quantitatively measured by the labeling-error, defined using the entrywise Kronecker delta $\delta(\cdot)$, as $\|\delta(\boldsymbol{z}_0 - \hat{\boldsymbol{z}})\|_1/N_g$. Progression of the labeling error averaged over 20 Monte Carlo (MC) runs is displayed in Fig. 5.5a, where the proposed method consistently outperforms the non-adaptive one. This shows that informative measurements adaptively collected to decrease uncertainty of $\boldsymbol{f}$ given a current estimate of $\boldsymbol{\theta}$ improve

Table 5.1: Hyper-parameters of $\boldsymbol{\theta}_f$ for synthetic data tests.

m_1	m_2	m_3	m_4	σ_1^2	σ_2^2	σ_3^2	σ_4^2
0	0.9	2.7	5.3	10^{-4}	10^{-4}	10^{-4}	10^{-4}
a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4
0.8	0.8	0.8	0.8	1	1	0.5	0.5

accuracy of $\hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{z}}$ in the next time slot. As a result, the SLF reconstruction accuracy of Alg. 9 improves accordingly with fewer measurements, as confirmed by comparing Figs. 5.4c and 5.4k.

The next experiment tests robustness of the proposed algorithms against measurement noise ν_τ. We adopted the labeling-error for $\boldsymbol{z}$ averaged over sensor locations and realizations of $\{\nu_\tau\}_\tau^t$ to quantify the reconstruction performance. Fig. 5.5b shows the progression of the labeling error at $\tau = 8$ as a function of the noise precision φ_ν averaged over 20 MC runs. Note that Figs. 5.4d and 5.4f correspond to the rightmost point of the x-axis of Fig. 5.5b. Clearly, the reconstruction performance does not severely decrease as φ_ν decreases, or equivalently σ_ν^2 increases. This confirms that the proposed algorithm is reasonably robust against measurement noise.

Averaged estimates of $\boldsymbol{\theta}$ and associated standard deviation denoted with $\pm$ are listed in Table 5.2. Together with Fig. 5.4, the high estimation accuracy of hyperparameters implies that the proposed method can effectively reveal patterns of objects in $\mathcal{A}$ by correctly inferring the underlying statistical properties of each piecewise homogeneous region in the SLF. Note that $\boldsymbol{\varphi}_f$ entries are overestimated in Table 5.2. This can be intuitively understood in the sense that minimizing the KL divergence in (5.14) leads to $q(\boldsymbol{\theta}_f)$ avoiding regions in which $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{\theta}_f)$ is small by setting each φ_{f_k} to a large value $\forall k$, which corroborates the result in [10, p. 468].

Next, we will validate the efficacy of Alg. 9 for channel estimation. Since g_0 and γ are known, obtaining $s(\mathbf{x}, \mathbf{x}')$ is equivalent to finding $g(\mathbf{x}, \mathbf{x}')$; cf. (2.1). This suggests adopting a performance metric quantifying the mismatch between $s(\mathbf{x}, \mathbf{x}')$ and $\hat{s}(\mathbf{x}, \mathbf{x}')$, using the normalized mean-square error

$$\text{NMSE} := \frac{\mathbb{E}\left[\int_{\mathcal{A}} \left(s(\mathbf{x}, \mathbf{x}') - \hat{s}(\mathbf{x}, \mathbf{x}')\right)^2 d\mathbf{x} d\mathbf{x}'\right]}{\mathbb{E}\left[\int_{\mathcal{A}} s^2(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}'\right]} \tag{5.47}$$

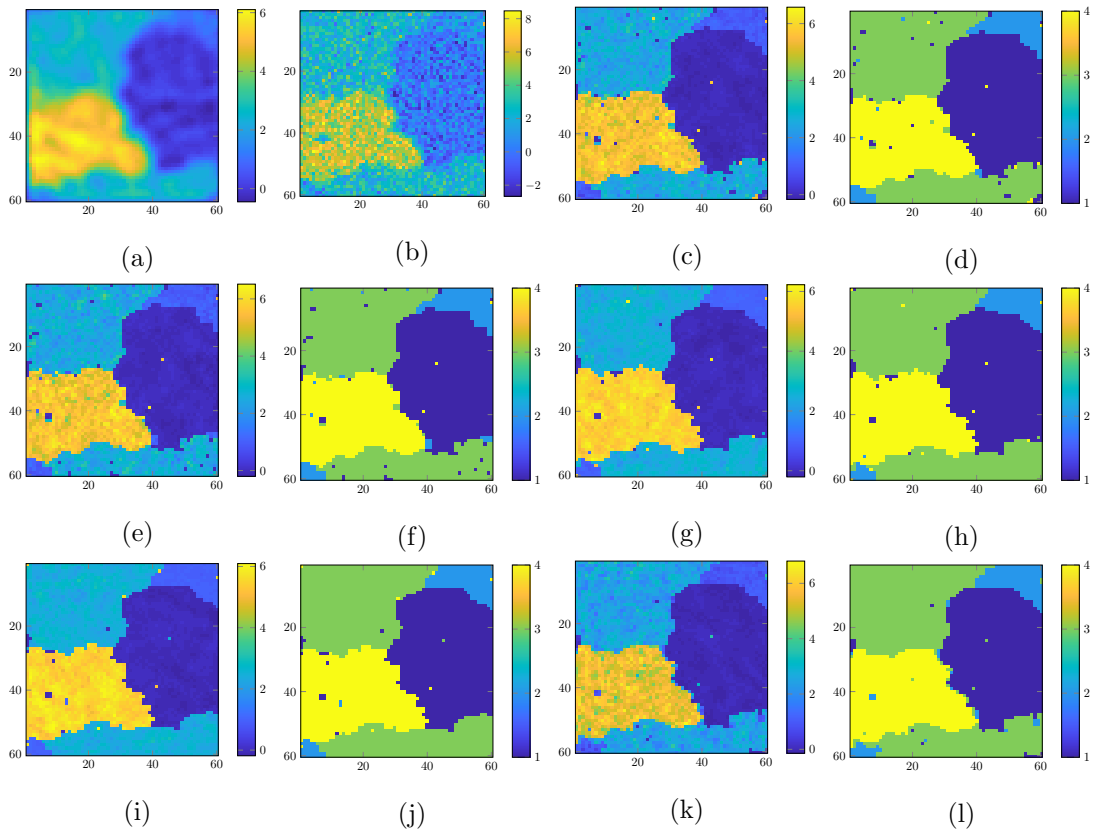
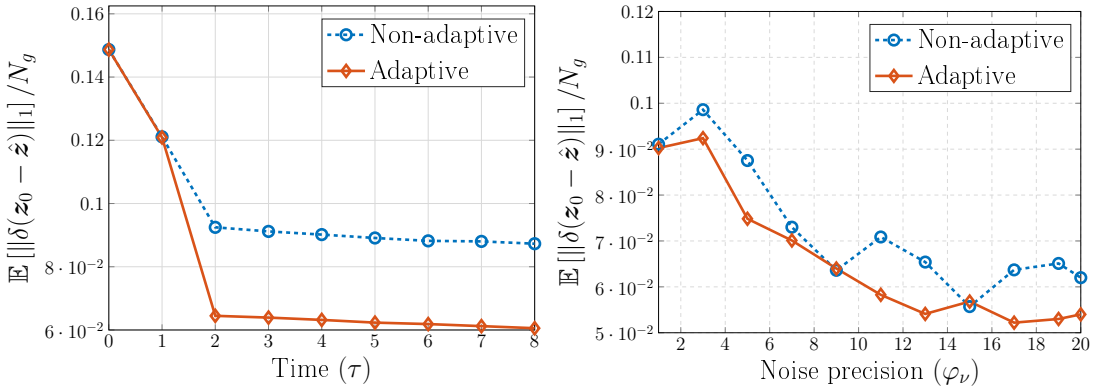


Figure 5.4: SLF estimates $\widehat{\mathbf{F}}$ at $\tau = 8$ (with $1,600$ measurements) via; (a) ridge-regularized LS ($\mu_f = 0.015$ and $\mathbf{C}_f = \mathbf{I}_{3,600}$); (b) TV-regularized LS ($\mu_f = 10^{-11}$); (c) Alg. 9 through (d) estimated hidden field $\widehat{\mathbf{Z}}$; (e) non-adaptive VB algorithm through (f) $\widehat{\mathbf{Z}}$; (g) adaptive MCMC algorithm through (h) $\widehat{\mathbf{Z}}$; (i) non-adaptive MCMC algorithm through (j) $\widehat{\mathbf{Z}}$; and (k) $\widehat{\mathbf{F}}_{\text{full}}$ and (l) $\widehat{\mathbf{Z}}_{\text{full}}$ obtained by using the full data (with $2,400$ measurements) via Alg. 9.

where the expectation is over the set $\{\mathbf{x}_n\}_{n=1}^N$ of sensor locations and realizations of $\{\nu_\tau\}_\tau$. The integrals are approximated by averaging the integrand over 500 pairs of $(\mathbf{x}, \mathbf{x}')$ chosen independently and uniformly at random on the boundary of $\mathcal{A}$. The expectations are estimated by averaging simulated deviates over 20 MC runs.

Fig. 5.6 depicts the NMSE of the proposed method and those of competing alternatives. Clearly, the approximate Bayesian inference methods outperform the regularized LS solutions. Furthermore, the performance of the VB methods is comparable to those of the MCMC methods. Noticeably, the adaptive VB method consistently exhibits



(a) (b)

Figure 5.5: Progression of estimation error of $\boldsymbol{z}$ versus (a) time τ; and (b) noise precision φ_ν, averaged over 20 MC runs.

Table 5.2: True $\boldsymbol{\theta}$ and estimated $\hat{\boldsymbol{\theta}}$ via Alg. 9 (setting of Fig. 5.4c); and non-adaptive VB algorithm (setting of Fig. 5.4e) averaged over 20 independent MC runs.

$\boldsymbol{\theta}$	True	Est. (Alg. 9)	Est. (non-adaptive)
φ_ν	20	$18.329 \pm 6 \times 10^{-3}$	$18.461 \pm 4.6 \times 10^{-3}$
μ_{f_1}	0	$0.022 \pm 1.2 \times 10^{-2}$	$0.018 \pm 1.9 \times 10^{-2}$
μ_{f_2}	1	$0.957 \pm 1.7 \times 10^{-2}$	$0.962 \pm 1.6 \times 10^{-2}$
μ_{f_3}	2.5	$2.573 \pm 1.7 \times 10^{-2}$	$2.578 \pm 2.6 \times 10^{-2}$
μ_{f_4}	5.5	$5.399 \pm 2.7 \times 10^{-2}$	$5.374 \pm 7 \times 10^{-3}$
φ_{f_1}	10	$40.178 \pm 3 \times 10^{-3}$	$42.352 \pm 2 \times 10^{-3}$
φ_{f_2}	10	$14.634 \pm 1.4 \times 10^{-2}$	$15.845 \pm 1.2 \times 10^{-2}$
φ_{f_3}	2	$7.712 \pm 2.7 \times 10^{-2}$	$7.493 \pm 2.2 \times 10^{-2}$
φ_{f_4}	2	$4.620 \pm 6.1 \times 10^{-2}$	$5.451 \pm 4 \times 10^{-2}$

lower NMSE than both non-adaptive ones, which highlights the efficacy in estimating channel-gain via the data-adaptive sensor selection. This suggests that the proposed VB framework is a viable solution for channel-gain cartography, while enjoying low computational complexity.

5.5.2 Test with real data

This section validates the proposed method using the real dataset in [33]. The test setup is depicted in Fig. 3.6, where $\mathcal{A} = [0.5, 20.5] \times [0.5, 20.5]$ is a square with sides of

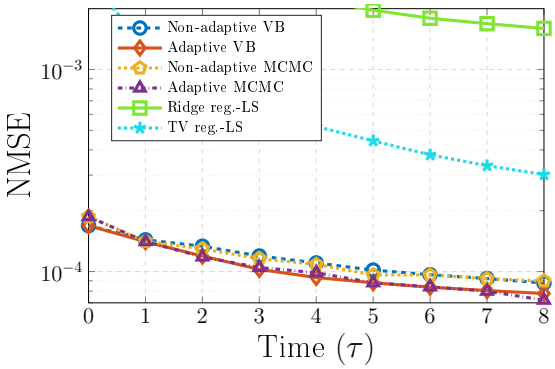


Figure 5.6: Progression of channel-gain estimation error.

20 feet (ft), over which a grid $\{\tilde{\mathbf{x}}_i\}_{i=1}^{3,721} := \{1,\ldots,61\}^2$ of $N_g = 3,721$ points is defined. A collection of $N = 80$ sensors measure the RSS at 2.425 GHz between pairs of sensor positions, marked with the $N = 80$ crosses in Fig. 3.6. To estimate g_0 and γ using the approach in [33], a first set of $2,400$ measurements was obtained before placing objects. Estimates $\hat{g}_0 = 54.6$ (dB) and $\hat{\gamma} = 0.276$ were obtained during the calibration phase. Afterwards, the structure comprising one pillar and six walls of different materials was assembled as shown in Fig. 3.6, and $T = 2,380$ measurements $\{\check{g}_{\tau'}\}_{\tau'=1}^{T}$ were collected. Calibrated measurements $\{\check{s}_{\tau'}\}_{\tau'=1}^{T}$ were then obtained from $\{\check{g}_{\tau'}\}_{\tau'=1}^{T}$ after substituting $\hat{g}_0$ and $\hat{\gamma}$ into (2.5). The weights $\{\mathbf{w}_{\tau'}^{(n,n')}\}_{\tau'=1}^{T}$ were constructed with w in (2.3) by using known locations of sensor pairs. Note that τ' is introduced to distinguish indices of the real data from τ used to index time slots in numerical tests.

We randomly selected $1,380$ measurements from $\{\check{s}_{\tau'}\}_{\tau'=1}^{T}$ to initialize $\check{\mathbf{s}}^{(0)}$ and $\mathbf{W}^{(0)}$, and used the remaining $1,000$ measurements to run the proposed algorithm under the mini-batch operation for $\tau = 0,1,\ldots,5$. At each time slot τ, $\mathcal{M}_{\tau+1}$ was formed by sensors corresponding to $|\mathcal{M}_{\tau+1}| = 200$ weight vectors uniformly selected at random from $\{\mathbf{w}_{\tau'}^{(n,n')}\}_{\tau'}$ associated with the remaining $1,000$ measurements without replacement. Then, $N_{\text{Batch}} = 100$ measurements were chosen from $\{\check{s}_{\tau'}\}_{\tau'}$ associated with $\mathcal{M}_{\tau+1}$.

Simulation parameters were set to $N_{\text{Iter}} = 3,000$, $\xi = 10^{-6}$, and $K = 3$; and hyper-hyper parameters of $\boldsymbol{\theta}$ were set to $a_\nu = b_\nu = 10^{-3}$, $[m_1, m_2, m_3]^\top = [0, 0.035, 0.05]^\top$, $\sigma_k^2 = 10^{-4}$ $\forall k$, and $a_k = b_k = 0.1$ $\forall k$, respectively. To execute Alg. 8, variational parameters of $q^{(0)}(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})$ were initialized as follows: $\left\{\breve{\mu}_{f_k}^{(0)}(\tilde{\mathbf{x}}_i)\right\}_{i=1}^{N_g}$ $\forall k$, $\breve{b}_\nu^{(0)}$, $\left\{\breve{\sigma}_k^{2(0)}\right\}_{k=1}^{3}$,

and $\left\{\breve{a}_k^{(0)}, \breve{b}_k^{(0)}\right\}_{k=1}^3$ were drawn from the uniform distribution $\mathcal{U}(0,1)$, while $\breve{m}_k^{(0)} = m_k \ \forall k$ and $\breve{\zeta}_k^{(0)}(\tilde{\mathbf{x}}_i) = 1/3 \ \forall i,k$.

Following [2, 33], a spatial covariance matrix was used for $\mathbf{C}_f$ of the ridge-regularized LS estimator, which models the similarity between points $\tilde{\mathbf{x}}_i$, and $\tilde{\mathbf{x}}_j$ as $\left[\mathbf{C}_f\right]_{ij} = \sigma_s^2 \exp[-\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2/\kappa]$ [2] with $\sigma_s^2 = \kappa = 1$, and $\mu_f = 0.015$ found with the L-curve [49, Chapter 26]. For the TV-regularized LS estimator, it was set to $\mu_f = 6$ found through the generalized cross validation [31]. To assess the efficacy of our Bayesian model with the K-ary hidden label field, we tested the adaptive MCMC method in [50] with $K = 2$.

Figs. 5.7a–5.7h depict SLF estimates $\widehat{\mathbf{F}}$ and associated hidden fields $\widehat{\mathbf{Z}}$ at $\tau = 5$ obtained via the proposed algorithms and competing alternatives. As a benchmark, one-shot estimates of the SLF $\widehat{\mathbf{F}}_{\text{full}}$ and associated hidden field $\widehat{\mathbf{Z}}_{\text{full}}$ are also displayed in Figs. 5.7i and 5.7j obtained via Alg. 9 by using the entire set of $2,380$ measurements. Comparing Figs. 5.7e and 5.7i (or Figs. 5.7f and 5.7j) shows that the proposed method accurately reveals the structural pattern of the testbed by using fewer number of measurements; e.g., the cinder block in the testbed was not captured by the SLF in Fig. 5.7g, but that in Fig. 5.7e. For competing alternatives, the testbed structure was not captured through the SLFs in Figs. 5.7a and 5.7b estimated via both regularized LS methods. On the other hand, the MCMC method reveals the structure through $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{Z}}$ in Figs. 5.7c and 5.7d, although they are less accurately delineated than those from the proposed method. This illustrates the benefits of considering a general Bayesian model with $K \geq 2$ addressing a richer class of spatial heterogeneity.

Efficacy of the data-driven sensor selection scheme is further analyzed. Specifically, the accuracy of $\hat{\boldsymbol{z}}$ measured by the labeling error $\|\delta(\hat{\boldsymbol{z}}_{\text{full}} - \hat{\boldsymbol{z}})\|_1/N_g$ with $\hat{\boldsymbol{z}}_{\text{full}} := \text{vec}(\widehat{\mathbf{Z}}_{\text{full}})$ was used as performance metric. Progression of the labeling error for Alg. 9 is depicted in Fig. 5.8 with that for the non-adaptive VB algorithm, where the proposed method consistently outperforms the non-adaptive one for every τ. This implies that the proposed sensor selection strategy helps to reveal object patterns more accurately while reducing data collection costs.

To corroborate the hyperparameter estimation capability of the proposed algorithm, estimates of $\boldsymbol{\theta}$ averaged over 20 MC runs are listed in Table 5.3. Estimated $\hat{\boldsymbol{\theta}}$ obtained by using the full data was considered as a benchmark, to demonstrate that Alg. 9 yields estimates $\boldsymbol{\theta}$ closer to the benchmark than its non-adaptive counterpart (except φ_ν).

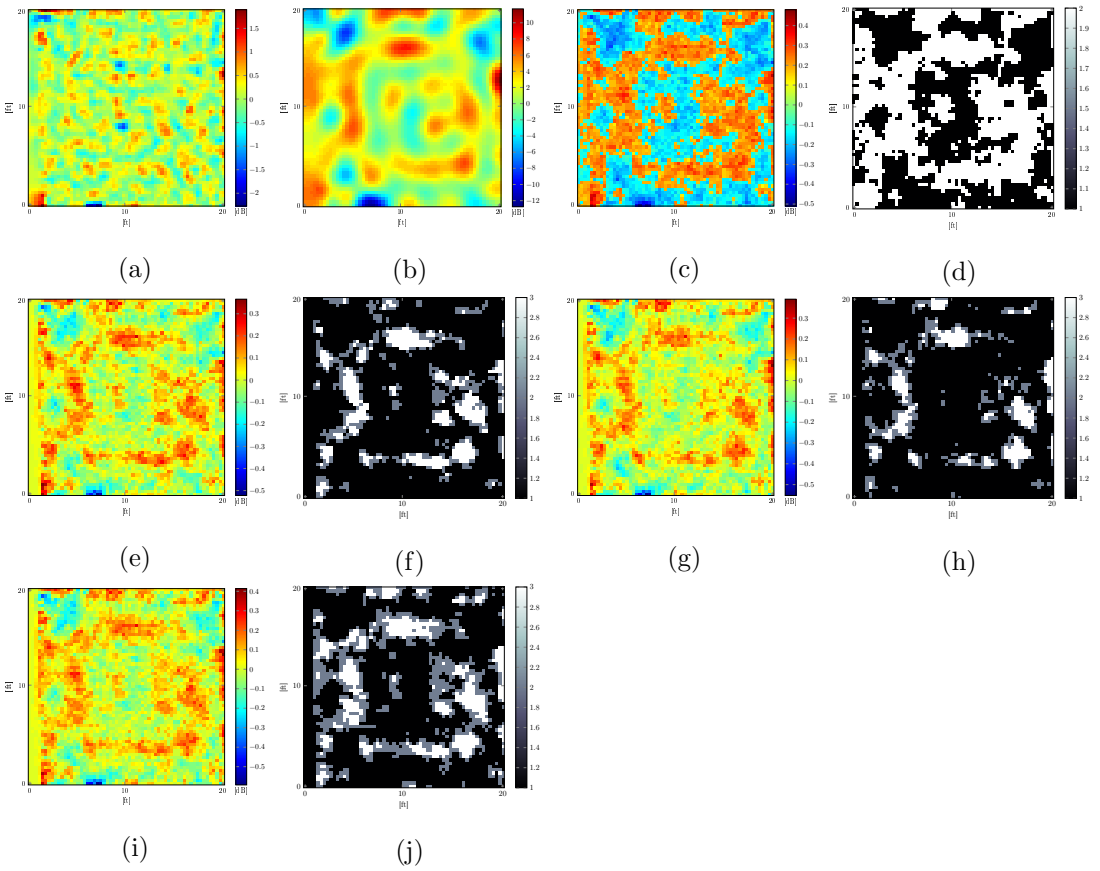


Figure 5.7: SLF estimates $\widehat{\mathbf{F}}$ at $\tau = 5$ (with $1,880$ measurements) via; (a) ridge-regularized LS ($\mu_f = 0.015$ and $\mathbf{C}_f = \mathbf{I}_{3,600}$); (b) TV-regularized LS ($\mu_f = 6$); (c) adaptive MCMC algorithm in [50] with $K = 2$ through (d) estimated hidden field $\widehat{\mathbf{Z}}$; (e) Alg. 9 through (f) $\widehat{\mathbf{Z}}$; (g) non-adaptive VB algorithm through (h) $\widehat{\mathbf{Z}}$; and (i) $\widehat{\mathbf{F}}_{\text{full}}$ and (j) $\widehat{\mathbf{Z}}_{\text{full}}$ obtained by using the full data (with $2,380$ measurements).

Note that the level of measurement noise is high since $\widehat{\sigma_\nu^2} = \widehat{\varphi_\nu}^{-1} \approx 15$. This can be justified because the testbed structure was accurately revealed in $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{Z}}$ from the proposed method by incorporating imperfect calibration effects in the measurement noise.

The last simulation assesses performance of the proposed algorithms for channel-gain map construction. The set of shadowing measurements and setup was the one used in the first simulated tests of this section. A channel-gain map is constructed to portray the

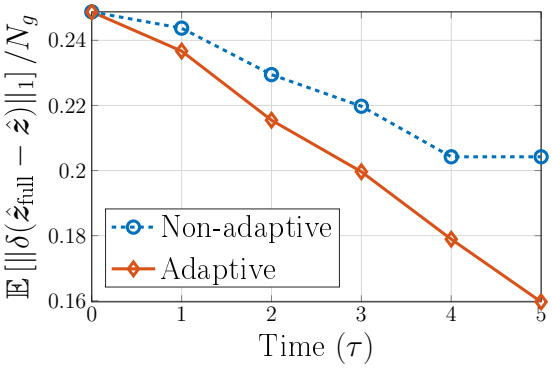


Figure 5.8: Progression of a mismatch between $\hat{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}_{\text{full}}$.

channel-gain between every point in the map $\mathbf{x}$, and a fixed receiver location $\mathbf{x}_{\text{rx}}$. Specifically, Alg. 9 is executed and estimates $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\hat{\boldsymbol{f}}$ and w into (2.4). Subsequently, $\{\hat{g}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ into (2.1) with $\hat{g}_0$ and $\hat{\gamma}$. Upon defining $\hat{\boldsymbol{g}} := [\hat{g}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \ldots, \hat{g}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top \in \mathbb{R}^{N_g}$, we construct the channel-gain map $\widehat{\mathbf{G}} := \text{unvec}(\hat{\boldsymbol{g}})$ with the receiver located at $\mathbf{x}_{\text{rx}}$.

Let $\widehat{\mathbf{S}} := \text{unvec}(\hat{\boldsymbol{s}})$ denote a shadowing map with $\hat{\boldsymbol{s}} := [\hat{s}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \ldots, \hat{s}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top \in \mathbb{R}^{N_g}$. Fig. 5.9 displays estimated shadowing maps and corresponding channel-gain maps constructed via Alg. 9 and the competing alternatives, when the receiver is located at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked by the cross. In every channel-gain map of Fig. 5.9, stronger attenuation is observed when signals propagate through either more building materials (bottom-right side of $\widehat{\mathbf{G}}$), or the concrete wall (left side of $\widehat{\mathbf{G}}$). On the other hand, only the channel-gain maps in Figs. 5.9f, 5.9h, 5.9j, and 5.9l constructed by the approximate Bayesian inference methods exhibit less attenuation along the entrance of the structure (top-right side of $\widehat{\mathbf{G}}$); this cannot be seen through the channel-gain maps in Figs. 5.7a and 5.7b constructed by both regularized LS methods. The reason is that free space and objects are more distinctively delineated in $\widehat{\mathbf{F}}$ by the proposed method. All in all, the simulation results confirm that our approach could provide more site-specific information of the propagation medium, and thus endows the operation of cognitive radio networks with more accurate interference management.

Table 5.3: Estimated $\hat{\boldsymbol{\theta}}$ via benchmark algorithm (setting of Fig. 5.7i); Alg. 9 (setting of Fig. 5.7e); and non-adaptive VB algorithm (setting of Fig. 5.7g), averaged over 20 independent MC runs.

$\boldsymbol{\theta}$	Benchmark	Est. (Alg. 9)	Est. (non-adaptive)
φ_ν	0.075 ± 10^{-16}	0.068 ± 0.13	0.071 ± 0.24
μ_{f_1}	-0.001 ± 10^{-17}	-0.001 ± 10^{-7}	-0.001 ± 10^{-7}
μ_{f_2}	0.032 ± 10^{-17}	0.032 ± 10^{-8}	0.032 ± 10^{-8}
μ_{f_3}	0.045 ± 10^{-17}	0.046 ± 10^{-8}	0.046 ± 10^{-8}
φ_{f_1}	5.524 ± 10^{-18}	4.951 ± 10^{-3}	$4.789 \pm 1.9 \times 10^{-3}$
φ_{f_2}	5.524 ± 10^{-18}	4.942 ± 10^{-3}	$4.782 \pm 1.7 \times 10^{-3}$
φ_{f_3}	5.524 ± 10^{-18}	4.935 ± 10^{-3}	$4.775 \pm 1.7 \times 10^{-3}$

5.6 Conclusion

This section developed a variational Bayes approach to adaptive channel-gain cartography, which estimates the spatial loss field of the tomographic model at affordable complexity by using measurements collected from sensing radio pairs that are adaptively chosen with an uncertainty sampling criterion. Extensive synthetic and real data tests corroborated the efficacy of the proposed novel algorithm for channel-gain cartography and tomographic imaging applications.

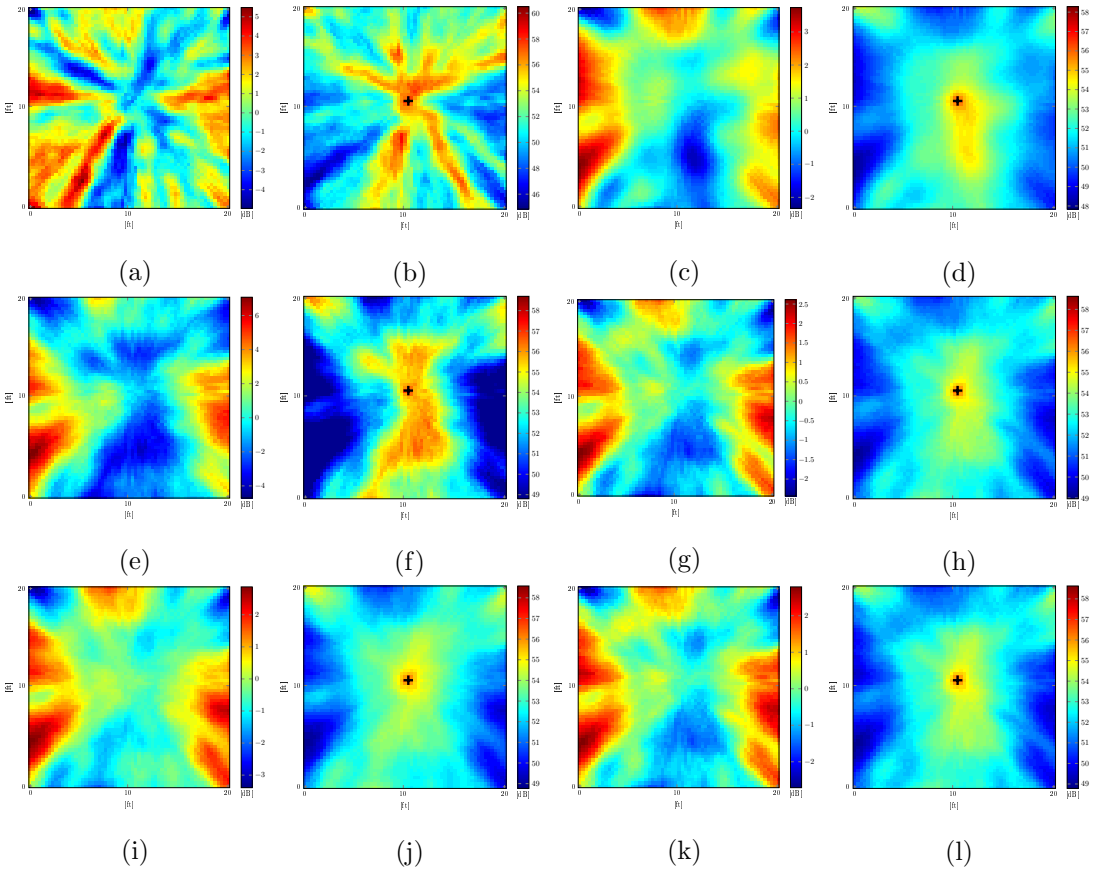


Figure 5.9: Estimated shadowing maps $\widehat{\mathbf{S}}$ and corresponding channel-gain maps $\widehat{\mathbf{G}}$ at $\tau = 5$ via (a)-(b) ridge-regularized LS (setting of Fig. 5.7a); (c)-(d) TV-regularized LS (setting of Fig. 5.7b); (e)-(f) adaptive MCMC algorithm in [50] with $K = 2$ (setting of Fig. 5.7c); (g)-(h) Alg. 9 (setting of Fig. 5.7e); and (i)-(j) non-adaptive VB algorithm (setting of Fig. 5.7g); and (k)–(l) benchmark algorithm (setting of Fig. 5.7i), with the receiver location at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked with the black cross.

Chapter 6

Summary and Future Directions

Leveraging recent advances in statistical signal processing and machine learning, this thesis contributed novel algorithms that help realize the goal of data-driven channel learning to design and operate next-generation wireless communication networks. The following subsections provide a summary of the work presented in this thesis, as well as possible future research directions.

6.1 Thesis Summary

After the review on channel-gain cartography in Chapter 2, a low-rank plus sparse matrix model was presented for channel-gain cartography in Chapter 3, which is instrumental for spectrum sensing and resource allocation tasks to operate cognitive radio networks. Channel-gain was modeled as the aggregate effect (in dB) of distance-based pathloss and shadowing expressed as tomographic accumulation of the underlying spatial loss field (SLF). The SLF was postulated to have a low-rank structure corrupted by sparse outliers. Efficient batch and online algorithms were developed by leveraging a bilinear characterization of the matrix nuclear norm. The algorithms enjoy low computational complexity and a reduced memory requirement, without sacrificing optimality, and with provable convergence properties. Tests with both synthetic and real datasets corroborated the claims and showed that the algorithms could accurately reveal the structure of propagation medium.

Chapter 4 dealt with the development of a novel Bayesian framework for adaptive

channel-gain cartography that estimates the SLF by using measurements collected by a set of sensing radio pairs chosen based on the uncertainty sampling criterion. Different from conventional approaches, leveraging a binary hidden label field accounted for heterogeneity of the SLF. The effectiveness of the novel Markov chain Monte Carlo (MCMC) algorithm was corroborated through extensive synthetic and real data experiments.

Finally, Chapter 5 was built upon the Bayesian framework and the results of Chapter 4. The proposed Bayesian SLF model was further generalized to have K-ary piecewise homogeneous regions, which accounts for a richer class of environmental heterogeneity. As an alternative to the computationally expensive MCMC algorithms, we developed the variational Bayes (VB) solution for channel-gain cartography, which estimates the SLF at affordable complexity by using measurements acquired by sensing radio pairs chosen by another data-adaptive sensor selection strategy. Extensive synthetic and real data tests corroborated the efficacy of the proposed novel algorithm for channel-gain cartography and radio tomographic imaging applications.

6.2 Future Research

The promising results in this thesis open up interesting directions for a number of future research topics. The following subsections discuss a few of these directions.

6.2.1 Online Bayesian channel-gain cartography

The proposed method in Chapter 5 outperformed conventional solutions for channel-gain cartography and radio tomography. When it comes to the scenario of streaming data however, the current setup is not fully suitable since a proxy of the posterior $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)$ needs to be updated in a batch fashion every time a new datum $\check{s}_{t+1}$ arrives. This issue is frequently encountered particularly in large-scale learning via approximate Bayesian inference. To bypass this limitation while efficiently coping with streaming datasets, we propose using *online* approximate Bayesian inference algorithms. Examples of such algorithms include the stochastic variational inference [39] and the streaming variational Bayes [13], which updates variational parameters in the spirit of stochastic approximation (SA) [48, 84] utilizing natural gradients of the evidence lower bound.

6.2.2 Variational massive MIMO channel estimation

Hybrid analog and digital beamforming (HBF) with low-resolution ADCs (1–3 bits) is a key enabler of mmWave communication via massive MIMO techniques, by addressing the associated hardware issues of costly implementation and power consumption of devices. Designing the HBF architecture requires knowing the channel-state information to configure the precoder and combiner in analog and digital domains. Unfortunately, it is challenging to obtain the channel information in a massive MIMO system with HBF for several reasons: i) the channel matrix has large dimension; ii) low receive-SNR; and iii) indirect access to the channel matrix. These factors prevent one from using channel estimation techniques for conventional beamforming. These considerations motivate developing tools for scalable, yet accurate channel estimation in accordance with the pressing needs of the HBF structure.

By leveraging the sparse nature of mmWave channel [3, 4, 37], a large body of works casts the channel estimation problem as an instance of *noisy quantized compressed sensing*. To tackle the latter, approximate message passing (AMP)-based methods, including generalized AMP (GAMP) [76] and vector AMP (VAMP) [78], have been widely employed. AMP methods were originally developed to efficiently solve the basis pursuit or LASSO problem [21], and provide MMSE solutions for channel estimation in the large system limit with a regressor matrix having i.i.d. (sub-) Gaussian entries. Otherwise, AMP-based algorithms trade-off convergence guarantees for computational complexity. To bypass such limitations, one can resort to approximate Bayesian inference techniques, including variational Bayes expectation-maximization (VBEM) methods, whose convergence does not suffer from (sub-)Gaussianity of the regressor matrix. To leverage the sparsity present in the mmWave channel, the posterior model can be designed with proper sparsity inducing priors such as i) the Bernoulli-Gaussian (BG); or the ii) Bernoulli Gaussian-mixture (BGM) oner [63].

6.2.3 Angular pattern reconstruction of mmWave channel

Directional transmissions and MIMO techniques with adaptive beamforming gain attentions as a means of compensating severe signal attenuation at mmWave bands. To efficiently construct a beam steering vector, it is essential to know angle-of-arrivals

and angular-power distribution. Conventionally, empirical approaches based on field measurements have been used to model mmWave propagation [79, 83, 77] to analyze path-loss or to estimate angle-of-arrivals. Unfortunately, the resulting performance might be limited since angular resolution of the antenna highly depends on its effective beamwidth, which leads to angular spread estimation under simplified assumption. Otherwise, data collection might be too expensive due to necessity of the antenna with highly directive beam such as the pencil-beam. To overcome aforementioned limitation, we propose a data-driven approach to infer angular response of channel in high resolution, given a non-coherent power-angular measurement at a receiver with its antenna radiation pattern. Therefore, the reconstructed angular pattern of channel can provide sufficient granularity to accurately reveal angle-of-arrivals, or to localize users, depending on applications of interest.

One can formulate this task as a instance of the *phase retrieval* (PR) problem. Consider an antenna measuring the received signal amplitude at every Azimuth angle $\phi_n \in [0, 2\pi)$ for $n = 1, \ldots, 360$, with unit angular resolution (in degrees). Let $\mathbf{a} \in \mathbb{C}^{360}$ denote the known antenna pattern with the n-th element $a(\phi_n) := a_n = |a_n| \exp(j\theta_{a_n})$, where $|a_n|^2$ and θ_{a_n} are the n-th antenna power and phase, respectively. With these notational conventions, the antenna measurement can be expressed as [74]

$$\boldsymbol{y} = |\mathbf{a} \circledast \boldsymbol{h} + \boldsymbol{\nu}| \tag{6.1}$$

where $\boldsymbol{h} \in \mathbb{C}^N$ is the channel in the angular domain with the n-th element $h(\phi_n) := h_n = |h_n| \exp(j\theta_{h_n})$; and $\boldsymbol{\nu} \in \mathbb{C}^N$ is noise. Given $\boldsymbol{y}$, the goal is to estimate $\boldsymbol{h}$, which is precisely a PR problem. However, the performance of conventional solutions [15, 74, 88] can be limited because the global optimum is not always guaranteed due to the structure of $\mathbf{a}$. This limitation also leads to slow convergence. As an alternative of the existing PR solutions, one venue of future research is to investigate a Bayesian framework based on a hierarchical graphical model by utilizing approximate Bayesian inference techniques.

References

[1] "IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements-Part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications Amendment 3: Enhancements for very high throughput in the 60 Ghz band," *IEEE Std 802.11ad-2012*, pp. 1–628, Dec. 2012.

[2] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4024–4036, Aug. 2009.

[3] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Sig. Proc.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[4] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. 2015 Intl. Conf. on Acoust., Speech, and Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 2909–2913.

[5] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss-Markov-Potts models and variational Bayesian computation," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2265 – 2277, Sep. 2010.

[6] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-Lasso on splines for spectrum cartography," *IEEE Trans. Sig. Proc.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011.

[7] M. J. Beal, *Variational Algorithms for Approximate Bayesian Inference*. London, U.K: Gatsby Comput. Neurosci. Unit, Univ. College London, 2003.

[8] B. Beck, X. Ma, and R. Baxley, "Ultrawideband tomographic imaging in uncalibrated networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6474 – 6486, Sep. 2016.

[9] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2006. [Online]. Available: https://books.google.com/books?id=kTNoQgAACAAJ

[11] J. K. Blitzstein and J. Hwang, *Introduction to Probability*. Boca Raton, FL: USA:CRC, 2014.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[13] T. Broderick, N. Boyd, A. Wibisono, A. C. Wilson, and M. I. Jordan, "Streaming variational Bayes," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013.

[14] R. F. Brown, *A Topological Introduction to Nonlinear Analysis*. Boston, MA: Birkhäuser, 2004.

[15] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," *Ann. Stat.*, vol. 44, no. 5, pp. 2221–2251, Nov. 2016.

[16] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. of the ACM*, vol. 58, no. 3, article no. 11, May 2011.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: USA:Wiley, 1991.

[18] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, 2011.

[19] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov Chain Monte Carlo maximum likelihood," *IEEE Trans. Sig. Proc.*, vol. 8, no. 7, pp. 954 – 963, Jul. 1999.

[20] Y. Dong, R. Schreier, W. Yang, S. Korrapati, and A. Sheikholeslami, "29.2 A 235 mW CT 0-3 MASH ADC achieving-167dBFS/Hz NSD with 53 MHz BW," in *Proc. of IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2014, pp. 480–481.

[21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914 – 18 919, Nov. 2009.

[22] N. Dunford and J. T. Schwartz, *Linear Operators. Part 1: General Theory.* Newyork, NY: Interscience publishers, 1958.

[23] Ericsson, "Ericsson mobility report," Tech. Rep. Q1 update, Jan. 2019, https://www.ericsson.com/4a517b/assets/local/mobility-report/documents/2019/emr-q2-2019-update.pdf.

[24] A. Eriksson and A. van den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L_1 norm," in *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, Jun. 2010, pp. 771–778.

[25] FCC, "Spectrum policty task force," Tech. Rep. ET Docket 02-135, Nov. 2002.

[26] V. V. Fedorov, *Theory of Optimal Experiments.* New York, NY: Academic Press, 1972.

[27] B. Floyd, S. Reynolds, U. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, "SiGe bipolar transceiver circuits operating at 60 Ghz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156–167, Jan. 2005.

[28] B. E. Fristedt and L. F. Gray, *A Modern Approach to Probability Theory.* Boston, MA: Birkhäuser, 1997.

[29] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[30] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.

[31] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross- validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215 – 223, May 1979.

[32] A. Gonzalez-Ruiz and Y. Mostofi, "Cooperative robotic structure mapping using wireless measurements - a comparison of random and coordinated sampling patterns," *IEEE Sensors J.*, vol. 13, no. 7, pp. 2571 – 2580, Jul. 2013.

[33] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Propagation modeling for radio frequency tomography in wireless networks," *IEEE J. Sel. Topics Sig. Proc.*, vol. 8, no. 1, pp. 55–65, Feb. 2014.

[34] J. Hammersley and P. Clifford, "Markov field on finite graphs and lattices," 1971. [Online]. Available: http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf

[35] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY: Springer-Verlag, 1997.

[36] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[37] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Sig. Proc.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[38] D. Higdon, "Spatial applications of Markov chain Monte Carlo for Bayesian inference," Ph.D. dissertation, Dept. Stat., Univ. Washington, Seattle, WA, 1994.

[39] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *J. Machine Learning Res.*, vol. 14, pp. 1303–1347, May 2013.

[40] F. HU, B. CHEN, and K. ZHU, "Full spectrum sharing in cognitive radio networks toward 5g: A survey," *IEEE Access*, vol. 6, pp. 15 754 – 15 776, 2018.

[41] Intel. A guide to the internet of things infographic. https://www.intel.com/content/www/us/en/internet-of-things/infographics/guide-to-iot.html.

[42] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: a partially collapsed Gibbs sampler method," *IEEE Trans. Sig. Proc.*, vol. 60, no. 6, pp. 2727 – 2743, 2012.

[43] I. Kakalou, K. E. Psannis, P. Krawiec, and R. Badea, "Cognitive radio network and network service chaining toward 5G: Challenges and requirements," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 145–151, Nov. 2017.

[44] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the accuracy of radio tomographic imaging using channel diversity," in *Proc. of the 9th IEEE Int. Conf. on Mobile Ad hoc and Sensor Syst.*, Las Vegas, NV, Oct. 2012, pp. 254–262.

[45] M. A. Kanso and M. G. Rabbat, "Compressed RF tomography for wireless sensor networks: Centralized and decentralized approaches," in *Int. Conf. Distributed Comput. Sensor Syst.* Springer, 2009, pp. 173–186.

[46] S.-J. Kim, E. Dall'Anese, J. A. Bazerque, K. Rajawat, and G. B. Giannakis, "Advances in spectrum sensing and cross-layer design for cognitive radio networks," in *Academic Press Library in Signal Processing: Volume 2 Communications and Radar Signal Processing R. Chellapa and S. Theodoridis, Eds.* Academic Press, 2014, ch. 9, pp. 471–502.

[47] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE J. Sel. Topics Sig. Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.

[48] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications.* New York, NY: Springer-Verlag, 2003.

[49] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Philadelphia, PA: SIAM, 1974.

[50] D. Lee, D. Berberidis, and G. B. Giannakis, "Adaptive Bayesian radio tomography," *IEEE Trans. Sig. Proc.*, vol. 67, no. 8, pp. 1964 – 1977, Apr. 2019.

[51] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953 – 5966, Nov. 2017.

[52] F. Liang, "A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants," *J. Stat. Comp. Simul.*, vol. 80, no. 9, pp. 1007–1022, Sep. 2010.

[53] Z. Lin, A. Ganesh, J. Right, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," *UIUC Technical Report UILU-ENG-09-2214*, pp. 1–18, 2009.

[54] J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski, "Modeling human blockers in millimeter wave radio links," *ZTE Commun.*, vol. 10, no. 4, pp. 23–28, Dec. 2012.

[55] D. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590 – 604, 1992.

[56] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Machine Learning Res.*, vol. 11, pp. 19–60, Jan. 2010.

[57] M. Marcus and B. Pattan, "Millimeter wave propagation: Spectrum management implications," *IEEE Microw. Mag.*, vol. 6, no. 2, pp. 54–62, Jun. 2005.

[58] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Sig. Proc.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.

[59] ——, "Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies," *IEEE Trans. Info. Theory*, vol. 59, no. 8, pp. 5186–5205, Aug. 2013.

[60] ——, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Trans. Sig. Proc.*, vol. 63, no. 10, pp. 2663–2677, May 2015.

[61] C. A. McGrory, D. M. Titterington, R. Reeves, and A. N. Pettitt, "Variational Bayes for estimating the parameters of a hidden Potts model," *Stat. Comput.*, vol. 19, no. 3, pp. 329 – 340, 2009.

[62] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.

[63] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Sig. Proc.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[64] J. Møller, A. N. Pettitt, R. Reeves, and K. K. Berthelsen, "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, vol. 93, no. 2, pp. 451–458, 2006.

[65] A. Moshtaghpour, M. A. Akhaee, and M. Attarifar, "Obstacle mapping in wireless sensor networks via minimum number of measurements," *IET Sig. Proc.*, vol. 10, no. 3, pp. 237–246, 2016.

[66] Y. Mostofi, "Compressive cooperative sensing and mapping in mobile networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 12, pp. 1769–1784, Dec. 2011.

[67] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.

[68] I. Murray, Z. Ghahramani, and D. J. C. MacKay, "MCMC for doubly-intractable distributions," in *Proc. of the 22nd Annual Conf. on Uncertainty in Artificial Intell.*, Cambridge, MA, Jul. 2006.

[69] Y. Nesterov, *Introductory Lectures on Convex optimization : A Basic Course*. Boston, MA: Kluwer Academic, 2004.

[70] M. Opper and D. Saad, *Advanced mean field methods: Theory and practice*. Cambridge, MA: MIT Press, 2001.

[71] N. Patwari and P. Agrawal, "Effects of correlated shadowing: Connectivity, localization, and RF tomography," in *Int. Conf. Info. Process. Sensor Networks*, St. Louis, MO, Apr. 2008, pp. 82–93.

[72] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm," *IEEE Trans. Image Process.*, vol. 2, no. 6, pp. 2385 – 2397, Jun. 2013.

[73] Z. Qin, D. Goldfarb, and S. Ma, "An alternating direction method for total variation denoising," *Optim. Method Softw.*, vol. 30, no. 3, pp. 594–615, 2015.

[74] Q. Qu, Y. Zhang, Y. Eldar, and J. Wright, "Convolutional phase retrieval," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017.

[75] Qualcomm. What can we do with 5G NR spectrum sharing that isn't possible today? [Online]. Available: https://www.qualcomm.com/media/documents/files/new-3gpp-effort-on-nr-in-unlicensed-spectrum-expands-5g-to-new-areas.pdf

[76] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. 2011 IEEE Int. Symp. Info. Theory*, 2011, pp. 2168–2172.

[77] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[78] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. 2017 IEEE Int. Symp. Info. Theory*, 2017, pp. 1588–1592.

[79] T. S. Rappaport, S. Sun, R. MAYZUS, H. ZHAO, Y. AZAR, K. W. G. N. WONG, J. K. SCHULZ, M. SAMIMI, and J. F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May. 2013.

[80] M. Razaviyayn, M. Sanjabi, and Z. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver

design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, 2016.

[81] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[82] D. Romero, D. Lee, and G. B. Giannakis, "Blind radio tomography," *IEEE Trans. Sig. Proc.*, vol. 66, no. 8, pp. 2055 – 2069, Apr. 2018.

[83] M. Samimi, K. Wang, Y. Azar, G. N. Wong, R. Mayzus, H. Zhao, J. K. Schulz, S. Sun, J. F. Gutierrez, and T. S. Rappaport, "28 Ghz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in New York City," in *Proc. Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013.

[84] K. Slavakis, S.-J. Kim, G. Mateos, and G. B. Giannakis, "Stochastic approximation vis-a-vis online learning for big data analytics," *IEEE Sig. Proc. Mag.*, vol. 31, Nov. 2014.

[85] D. Smith and M. Smith, "Estimation of binary Markov random fields using Markov chain Monte Carlo," *J. Comput. Graph. Stats.*, vol. 15, no. 1, pp. 207 – 227, 2006.

[86] P. Tseng, "Convergence of block coordinate descent method for nondifferentiable minimization," *J. Optimiz. Theory Applicat.*, vol. 109, pp. 475–494, Jun. 2001.

[87] R. A. Valenzuela, "Ray tracing prediction of indoor radio propagation," in *Proc. IEEE 5th Int. Symp. on Personal Indoor and Mobile Radio Comm.*, The Hague, Netherlands, Sep. 1994, pp. 140–144.

[88] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," *IEEE Trans. Sig. Proc.*, vol. 66, no. 2, pp. 479–491, Jan. 2018.

[89] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, 2010.

[90] ——, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 612–621, 2011.

[91] J. Wilson, N. Patwari, and O. G. Vasquez, "Regularization methods for radio tomographic imaging," in *Virginia Tech Symp. Wireless Personal Commun.*, Blacksburg, VA, Jun. 2009.

[92] WirelessHD. http://www.wirelesshd.org/.

[93] G. Wölfle, R. Wahl, P. Wertz, P. Wildbolz, and F. Landstorfer, "Dominant path prediction model for indoor scenarios," in *Proc. German Microwave Conf.*, 2005, pp. 176–179.

[94] K. Woyach, D. Puccinelli, and M. Haenggi, "Sensorless sensing in wireless networks: Implementation and measurements," in *Int. Symp. Modeling Optimization Mobile, Ad Hoc Wireless Netw.*, Apr. 2006, pp. 1–8.

[95] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, Jun. 2013.

[96] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 94 – 101, Apr. 2016.

[97] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. ACM MobiCom*, Sep. 2007, pp. 222–229.

[98] N. Zhao, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Joint segmentation and deconvolution of ultrasound images using a hierarchical Bayesian model based on generalized Gaussian priors," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3736 – 3750, 2016.

Appendix A

Proofs for Chapter 3

A.1 Proof of Proposition 1

A stationary point $\bar{\mathbf{P}}, \bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2) must satisfy the following first-order optimality conditions [12]

$$\mathbf{0}_{N_x \times N_y} \in \partial_{\mathbf{E}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \left\{ \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}}) + \mu_{\mathbf{E}} \bar{\beta} \left[\operatorname{sign}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}} \right] \middle| \bar{\mathbf{E}} \odot \tilde{\mathbf{E}} = \mathbf{0},\ \|\tilde{\mathbf{E}}\|_\infty \leq 1 \right\} \tag{A.1}$$

$$\nabla_{\mathbf{P}} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}})\bar{\mathbf{Q}} + \mu_{\mathbf{L}} \bar{\beta} \bar{\mathbf{P}} = \mathbf{0}_{N_x \times \rho} \tag{A.2}$$

$$\nabla_{\mathbf{Q}^\top} f(\bar{\mathbf{P}}, \bar{\mathbf{Q}}, \bar{\mathbf{E}}) = \bar{\mathbf{P}}^\top \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}}) + \mu_{\mathbf{L}} \bar{\beta} \bar{\mathbf{Q}}^\top = \mathbf{0}_{\rho \times N_y} \tag{A.3}$$

where $\odot$ denotes the element-wise (Hadamard) product. Through post-multiplying (A.2) by $\bar{\mathbf{P}}^\top$ and pre-multiplying (A.3) by $\bar{\mathbf{Q}}$, one can see that

$$\begin{aligned} \tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}}) &= -\mu_{\mathbf{E}} \bar{\beta} (\operatorname{sign}(\bar{\mathbf{E}}) + \tilde{\mathbf{E}}) \\ \operatorname{tr}\left(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top, \bar{\mathbf{E}}) \bar{\mathbf{Q}} \bar{\mathbf{P}}^\top \right) &= -\mu_{\mathbf{L}} \bar{\beta} \operatorname{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^\top) = -\mu_{\mathbf{L}} \bar{\beta} \operatorname{tr}(\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top). \end{aligned} \tag{A.4}$$

Define now $\kappa(\mathbf{R}_1, \mathbf{R}_2) := \frac{1}{2}\left(\mathrm{tr}(\mathbf{R}_1) + \mathrm{tr}(\mathbf{R}_2)\right)$, and consider the following *convex* problem

$$\text{(P4)} \quad \min_{\substack{\mathbf{L},\mathbf{E}\in\mathbb{R}^{N_x\times N_y},\\ \mathbf{R}_1\in\mathbb{R}^{N_x\times N_x},\\ \mathbf{R}_2\in\mathbb{R}^{N_y\times N_y}}} \sum_{\tau=1}^{T} \beta^{T-\tau} c^{(\tau)}(\mathbf{L},\mathbf{E}) + \mu_{\mathbf{L}}\bar{\beta}\ \kappa(\mathbf{R}_1,\mathbf{R}_2) + \mu_{\mathbf{E}}\bar{\beta}\ ||\mathbf{E}||_1$$

$$\text{subject to } \mathbf{R} := \begin{pmatrix} \mathbf{R}_1 & \mathbf{L} \\ \mathbf{L}^\top & \mathbf{R}_2 \end{pmatrix} \succeq \mathbf{0} \tag{A.5}$$

which is *equivalent* to (P1). Equivalence can be easily inferred by minimizing (P4) with respect to $\{\mathbf{R}_1, \mathbf{R}_2\}$ and noting an alternative characterization of the nuclear norm given by [81]

$$\begin{aligned} \|\mathbf{L}\|_* = \min_{\mathbf{R}_1,\mathbf{R}_2}\ & \kappa(\mathbf{R}_1,\mathbf{R}_2) \\ \text{subject to } & \mathbf{R} \succeq \mathbf{0}. \end{aligned} \tag{A.6}$$

In what follows, the optimality conditions of the conic program (P4) are explored. Introducing a Lagrange multiplier matrix $\mathbf{M} \in \mathbb{R}^{(N_x+N_y)\times(N_x+N_y)}$ associated with the conic constraint in (A.5), the Lagrangian is first formed as

$$\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \sum_{\tau=1}^{T} \beta^{T-\tau} c^{(\tau)}(\mathbf{L},\mathbf{E}) + \mu_{\mathbf{L}}\bar{\beta}\ \kappa(\mathbf{R}_1,\mathbf{R}_2) + \mu_{\mathbf{E}}\bar{\beta}\ \|\mathbf{E}\|_1 - \langle \mathbf{M}, \mathbf{R}\rangle. \tag{A.7}$$

For notational convenience, partition $\mathbf{M}$ as

$$\mathbf{M} := \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_4 & \mathbf{M}_3 \end{pmatrix} \tag{A.8}$$

in accordance with the block structure of $\mathbf{R}$ in (A.5), where $\mathbf{M}_1 \in \mathbb{R}^{N_x\times N_x}$ and $\mathbf{M}_3 \in \mathbb{R}^{N_y\times N_y}$. The optimal solution to (P4) must satisfy: (i) the stationarity conditions

$$\nabla_{\mathbf{L}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \tilde{f}(\mathbf{L},\mathbf{E}) - \mathbf{M}_2 - \mathbf{M}_4^\top = \mathbf{0} \tag{A.9}$$

$$\mathbf{0} \in \partial_{\mathbf{E}}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \left\{ \tilde{f}(\mathbf{L},\mathbf{E}) + \mu_{\mathbf{E}}\bar{\beta}\left[\mathrm{sign}(\mathbf{E}) + \tilde{\mathbf{E}}\right] \,\middle|\, \mathbf{E}\odot\tilde{\mathbf{E}} = \mathbf{0},\ \|\tilde{\mathbf{E}}\|_\infty \le 1 \right\} \tag{A.10}$$

$$\nabla_{\mathbf{R}_1}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\mathbf{I}_{N_x} - \mathbf{M}_1 = \mathbf{0} \tag{A.11}$$

$$\nabla_{\mathbf{R}_2}\mathcal{L}(\mathbf{L},\mathbf{E},\mathbf{R}_1,\mathbf{R}_2;\mathbf{M}) = \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\mathbf{I}_{N_y} - \mathbf{M}_3 = \mathbf{0} \tag{A.12}$$

(ii) complementary slackness condition $\langle \mathbf{M},\mathbf{R}\rangle = 0$; (iii) primal feasibility $\mathbf{R}\succeq\mathbf{0}$; and (iv) dual feasibility $\mathbf{M}\succeq\mathbf{0}$.

Using the stationary point $\bar{\mathbf{P}},\bar{\mathbf{Q}}$ and $\bar{\mathbf{E}}$ of (P2), construct a candidate solution for (P4) as $\widehat{\mathbf{L}} := \bar{\mathbf{P}}\bar{\mathbf{Q}}^\top$, $\widehat{\mathbf{E}} := \bar{\mathbf{E}}$, $\widehat{\mathbf{R}}_1 := \bar{\mathbf{P}}\bar{\mathbf{P}}^\top$, and $\widehat{\mathbf{R}}_2 := \bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top$, as well as $\widehat{\mathbf{M}}_1 := \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\mathbf{I}_{N_x}$, $\widehat{\mathbf{M}}_2 := \frac{1}{2}\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top,\bar{\mathbf{E}})$, $\widehat{\mathbf{M}}_3 := \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\mathbf{I}_{N_y}$, and $\widehat{\mathbf{M}}_4 := \widehat{\mathbf{M}}_2^\top$. After substituting these into (A.9)–(A.12), it can be readily verified that condition (i) holds. Condition (ii) also holds since

$$\begin{aligned}\langle \widehat{\mathbf{M}},\widehat{\mathbf{R}}\rangle &= \langle \widehat{\mathbf{M}}_1,\widehat{\mathbf{R}}_1\rangle + \langle \widehat{\mathbf{M}}_2,\widehat{\mathbf{L}}\rangle + \langle \widehat{\mathbf{M}}_3,\widehat{\mathbf{R}}_2\rangle + \langle \widehat{\mathbf{M}}_4,\widehat{\mathbf{L}}^\top\rangle \\ &= \frac{\mu_{\mathbf{L}}\bar{\beta}}{2}\mathrm{tr}(\bar{\mathbf{P}}\bar{\mathbf{P}}^\top + \bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top) + \mathrm{tr}\Big(\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top,\bar{\mathbf{E}})\bar{\mathbf{Q}}\bar{\mathbf{P}}^\top\Big) \\ &= \mathbf{0}\end{aligned} \tag{A.13}$$

where the last equality follows from (A.4). Condition (iii) is met since $\mathbf{R}$ can be rewritten as

$$\mathbf{R} = \begin{pmatrix} \bar{\mathbf{P}}\bar{\mathbf{P}}^\top & \bar{\mathbf{P}}\bar{\mathbf{Q}}^\top \\ \bar{\mathbf{Q}}\bar{\mathbf{P}}^\top & \bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{P}} \\ \bar{\mathbf{Q}} \end{pmatrix}\begin{pmatrix} \bar{\mathbf{P}} \\ \bar{\mathbf{Q}} \end{pmatrix}^\top \succeq \mathbf{0}. \tag{A.14}$$

For (iv), according to the Schur complement condition for positive semidefinite matrices, $\mathbf{M}\succeq\mathbf{0}$ holds if and only if

$$\widehat{\mathbf{M}}_3 - \widehat{\mathbf{M}}_4\widehat{\mathbf{M}}_1^{-1}\widehat{\mathbf{M}}_2 \succeq \mathbf{0} \tag{A.15}$$

which is equivalent to $\lambda_{\max}(\widehat{\mathbf{M}}_2^\top\widehat{\mathbf{M}}_2) \le (\mu_{\mathbf{L}}\bar{\beta}/2)^2$, or $||\tilde{f}(\bar{\mathbf{P}}\bar{\mathbf{Q}}^\top,\bar{\mathbf{E}})|| \le \mu_{\mathbf{L}}\bar{\beta}$. ■

A.2 Proof of Proposition 2

The proof uses the technique similar to the one employed in [56], where the convergence of online algorithms for optimizing objectives involving non-convex bilinear terms and

sparse matrices was established in the context of dictionary learning.

In order to proceed with the proof, three lemmata are first established. The first lemma concerns some properties of $g(\mathbf{X}, \boldsymbol{\xi}^{(t)}) := g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$, and $\check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) := \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$.

Lemma 2: *If the assumptions* (a1)–(a5) *in Proposition 2 hold, then*

(p1) $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ *majorizes* $g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$*, i.e.,* $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})\ \forall \mathbf{X} \in \mathcal{X}'$*;*

(p2) $\check{g}_1$ *is locally tight, i.e.,* $\check{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$*;*

(p3) $\nabla \check{g}_1(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$*;*

(p4) $\check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) := \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})$ *is uniformly strongly convex in* $\mathbf{X}$*, i.e.,* $\forall (\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \in \mathcal{X} \times \mathcal{X} \times \Xi$*, it holds that*

$$\check{g}(\mathbf{X} + \mathbf{D}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq \check{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + \frac{\zeta}{2}||\mathbf{D}||_F^2$$

where $\zeta > 0$ *is a constant and* $\check{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D})$ *is a directional derivative of* $\check{g}$ *at* $\mathbf{X}$ *along the direction* $\mathbf{D}$*;*

(p5) g_1 *and* $\check{g}_1$*, their derivatives, and their Hessians are uniformly bounded;*

(p6) g_2 *and its directional derivative* g_2' *are uniformly bounded; and*

(p7) *there exists* $\bar{g} \in \mathbb{R}$ *such that* $|\check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})| \leq \bar{g}$*.*

Proof: For (p1), let us first show that $\nabla_{\mathbf{P}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$, $\nabla_{\mathbf{Q}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$, and $\nabla_{\mathbf{E}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$ are Lipschitz continuous for $\mathbf{X} := (\mathbf{P}, \mathbf{Q}, \mathbf{E}) \in \mathcal{X}'$ and $\boldsymbol{\xi}^{(t)} \in \Xi$. For arbitrary $\mathbf{X}_1 := (\mathbf{P}_1, \mathbf{Q}_1, \mathbf{E}_1)$, $\mathbf{X}_2 := (\mathbf{P}_2, \mathbf{Q}_2, \mathbf{E}_2) \in \mathcal{X}'$, the variation of ∇g_1 in (3.19) can be bounded as

$$\begin{aligned}\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})\|_F &= \left\| \sum_{m=1}^{M} \langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1 - \mathbf{P}_2)\mathbf{Q}^\top \rangle \mathbf{W}_m^{(t)} \mathbf{Q} \right\|_F \\ &\overset{\text{(i1)}}{\leq} \sum_{m=1}^{M} |\langle \mathbf{W}_m^{(t)}, (\mathbf{P}_1 - \mathbf{P}_2)\mathbf{Q}^\top \rangle| \|\mathbf{W}_m^{(t)} \mathbf{Q}\|_F\end{aligned}$$

$$\overset{\text{(i2)}}{\leq} \sum_{m=1}^{M} \|\mathbf{P}_1 - \mathbf{P}_2\|_F \, \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2$$

where (i1) and (i2) are due to the triangle and Cauchy-Schwarz inequalities, respectively. Since $\mathcal{X}'$ and Ξ are assumed to be bounded, $\sum_{m=1}^{M} \|\mathbf{W}_m^{(t)}\mathbf{Q}\|_F^2$ is bounded. Therefore, there exists a positive constant $L_{\mathbf{P}}$ such that

$$\|\nabla_{\mathbf{P}} g_1(\mathbf{P}_1, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)}) - \nabla_{\mathbf{P}} g_1(\mathbf{P}_2, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})\|_F \leq L_{\mathbf{P}} \|\mathbf{P}_1 - \mathbf{P}_2\|_F \tag{A.16}$$

meaning that $\nabla_{\mathbf{P}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$ is Lipschitz continuous with constant $L_{\mathbf{P}}$. Similar arguments hold for $\nabla_{\mathbf{Q}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$ and $\nabla_{\mathbf{E}} g_1(\mathbf{P}, \mathbf{Q}, \mathbf{E}, \boldsymbol{\xi}^{(t)})$ as well, with Lipschitz constants $L_{\mathbf{Q}}$ and $L_{\mathbf{E}}$, respectively. Then, upon defining

$$\|\mathbf{X}\|_{\Delta} := \sqrt{L_{\mathbf{P}}^2 \|\mathbf{P}\|_F^2 + L_{\mathbf{Q}}^2 \|\mathbf{Q}\|_F^2 + L_{\mathbf{E}}^2 \|\mathbf{E}\|_F^2},$$

it is easy to verify

$$\|\nabla g_1(\mathbf{X}_1, \boldsymbol{\xi}^{(t)}) - \nabla g_1(\mathbf{X}_2, \boldsymbol{\xi}^{(t)})\|_F \leq \|\mathbf{X}_1 - \mathbf{X}_2\|_{\Delta}. \tag{A.17}$$

On the other hand, the proof of the Descent Lemma [9] can be adopted to show

$$g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) - g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \leq \langle \mathbf{X} - \mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \int_0^1 \|\mathbf{X} - \mathbf{X}^{(t-1)}\|_F \times \|\nabla g_1(\mathbf{X}^{(t-1)} + \alpha(\mathbf{X} - \mathbf{X}^{(t-1)}), \boldsymbol{\xi}^{(t)}) - \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})\|_F d\alpha. \tag{A.18}$$

Note that

$$\|\mathbf{X}\|_F \leq \frac{1}{L_{\min}} \|\mathbf{X}\|_{\Delta} \tag{A.19}$$

where $L_{\min} := \min\{L_{\mathbf{P}}, L_{\mathbf{Q}}, L_{\mathbf{E}}\}$. Then, substitution of (A.17) into (A.18) with $\mathbf{X}_1 = \mathbf{X}^{(t-1)} + \alpha(\mathbf{X} - \mathbf{X}^{(t-1)})$ and $\mathbf{X}_2 = \mathbf{X}^{(t-1)}$ yields

$$g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + \langle \mathbf{X} - \mathbf{X}^{(t-1)}, \nabla g_1(\mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \rangle + \frac{1}{2L_{\min}} \|\mathbf{X} - \mathbf{X}^{(t-1)}\|_{\Delta}^2 \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) \tag{A.20}$$

which completes the proof by the construction of $\check{g}_1$, provided that $\eta_i^{(t)} \geq L_i^2 / L_{\min}$ for

all $i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}$.

To show (p2) and (p3), let us first denote

$$\nabla g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) = \big(\nabla_{\mathbf{P}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}), \nabla_{\mathbf{Q}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}), \nabla_{\mathbf{E}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})\big) \tag{A.21}$$

$$\begin{aligned}\nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \big(&\nabla_{\mathbf{P}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{P}}^{(t)}(\mathbf{P} - \mathbf{P}^{(t-1)}),\\ &\nabla_{\mathbf{Q}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{Q}}^{(t)}(\mathbf{Q} - \mathbf{Q}^{(t-1)}),\\ &\nabla_{\mathbf{E}} g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)}) + \eta_{\mathbf{E}}^{(t)}(\mathbf{E} - \mathbf{E}^{(t-1)})\big).\end{aligned} \tag{A.22}$$

Then, it suffices to evaluate $\check{g}_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ and $\nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})$ at $\mathbf{X}^{(t-1)}$ to see that (p2) and (p3) hold.

To show (p4), let us first find $\check{g}_1'$ and g_2'. Along a direction $\mathbf{D} := (\mathbf{D}_{\mathbf{P}}, \mathbf{D}_{\mathbf{Q}}, \mathbf{D}_{\mathbf{E}}) \in \mathcal{X}'$, it holds that $\check{g}_1'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) = \langle \nabla \check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}), \mathbf{D} \rangle$ since $\check{g}_1$ is differentiable. Similarly, $g_2'(\mathbf{X}; \mathbf{D}) = \mu_{\mathbf{L}}(\langle \mathbf{P}, \mathbf{D}_{\mathbf{P}} \rangle + \langle \mathbf{Q}, \mathbf{D}_{\mathbf{Q}} \rangle) + \mu_{\mathbf{E}} h'(\mathbf{E}; \mathbf{D}_{\mathbf{E}})$ where $h(\mathbf{E}) := \|\mathbf{E}\|_1$, $\mathbf{d}_{\mathbf{E}} := \mathrm{vec}(\mathbf{D}_{\mathbf{E}})$ with its l-th entry being $d_{\mathbf{E},l}$, and

$$\begin{aligned} h'(\mathbf{E}; \mathbf{D}_{\mathbf{E}}) &:= \lim_{t \to 0+} \frac{h(\mathbf{E} + t\mathbf{D}_{\mathbf{E}}) - h(\mathbf{E})}{t} \\ &= \lim_{t \to 0+} \frac{\sum_{l, e_l \neq 0} (|e_l + t d_{\mathbf{E},l}| - |e_l|) + \sum_{l, e_l = 0} |t d_{\mathbf{E},l}|}{t} \\ &= \sum_{l, e_l \neq 0} \mathrm{sign}(e_l) d_{\mathbf{E},l} + \sum_{l, e_l = 0} |d_{\mathbf{E},l}|. \end{aligned} \tag{A.23}$$

On the other hand, the variation of $\check{g}$ can be written as

$$\begin{aligned}\check{g}(\mathbf{X} + \mathbf{D}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) = \check{g}_1'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + \sum_{i \in \{\mathbf{P}, \mathbf{Q}, \mathbf{E}\}} \frac{\eta_i^{(t)}}{2} \|\mathbf{D}_i\|_F^2 \\ + g_2(\mathbf{X} + \mathbf{D}) - g_2(\mathbf{X}).\end{aligned} \tag{A.24}$$

Note that $\sum_i \frac{\eta_i^{(t)}}{2} \|\mathbf{D}_i\|_F^2 \geq \frac{L_{\min}}{2} ||\mathbf{D}||_F^2$ since $\eta_i^{(t)} \geq L_i^2 / L_{\min}$ by algorithmic construction. Furthermore, $g_2(\mathbf{X} + \mathbf{D}) - g_2(\mathbf{X}) \geq g_2'(\mathbf{X}; \mathbf{D})$ since g_2 is convex [69]. Then, the variation of $\check{g}$ in (A.24) can be lower-bounded as

$$\check{g}(\mathbf{X} + \mathbf{D}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) - \check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq \check{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + \frac{L_{\min}}{2} ||\mathbf{D}||_F^2 \tag{A.25}$$

where $\check{g}'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) = \check{g}_1'(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}; \mathbf{D}) + g_2'(\mathbf{X}; \mathbf{D})$. Therefore, (p4) holds for a positive constant $\zeta \leq L_{\min}$.

By the compactness of $\mathcal{X}$ and boundedness of Ξ by (a3), (p5) is automatically satisfied since g_1 and $\check{g}_1$ are continuously twice differentiable in $\mathbf{X}$ [80]. In addition, one can easily show (p6) since g_2 and g_2' are also uniformly bounded by the compactness of $\mathcal{X}$.

Let K_1 and K_2 denote constants where $|\check{g}_1| \leq K_1$ and $|g_2| \leq K_2$, respectively, by (p5) and (p6). Then, (p7) readily follows since

$$\begin{aligned} |\check{g}(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})| &= |\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) + g_2(\mathbf{X})| \\ &\leq |\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)})| + |g_2(\mathbf{X})| \\ &\leq K_1 + K_2 =: \bar{g}. \blacksquare \end{aligned} \tag{A.26}$$

The next lemma asserts that a distance between two subsequent estimates asymptotically goes to zero, which will be used to show $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely.

Lemma 3: *If* (a2)–(a5) *hold, then* $||\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)}||_F = \mathcal{O}(1/t)$.

Proof: See [80, Lemma 2]. A proof of Lemma 3 is omitted to avoid duplication of the proof of [80, Lemma 2]. Hence, it suffices to mention that Lemma 2 guarantees the formulation of the proposed work satisfying the general assumptions on the formulation in [80]. ■

Lemma 3 does not guarantee convergence of the iterates to the stationary point of (P2). However, the final lemma asserts that the overestimated cost sequence converges to the cost of (P2), almost surely. Before proceeding to the next lemma, let us first define

$$C_{1,t}(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} g_1(\mathbf{X}, \boldsymbol{\xi}^{(\tau)}) \tag{A.27}$$

$$\check{C}_{1,t}(\mathbf{X}) := \frac{1}{t} \sum_{\tau=1}^{t} \check{g}_1(\mathbf{X}, \mathbf{X}^{(\tau-1)}, \boldsymbol{\xi}^{(\tau)}) \tag{A.28}$$

and $C_2(\mathbf{X}) := g_2(\mathbf{X})$. Note also that $\check{C}_t(\mathbf{X}) - C_t(\mathbf{X}) = \check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$.

Lemma 4: *If* (a1)–(a5) *hold,* $\check{C}_t(\mathbf{X}^{(t)})$ *converges almost surely, and* $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)})-$

$C_{1,t}(\mathbf{X}^{(t)}) = 0$, *almost surely.*

Proof: See [80, Lemma 1]. A proof of Lemma 4 is omitted to avoid duplication of the proof of [80, Lemma 1]. Instead, a sketch of the proof is following. It is firstly shown that the sequence $\{\check{C}_t(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ follows a quasi-martingale process and converges almost surely. Then, the lemma on positive converging sums (see [56, Lemma 8]) and Lemma 3 are used to claim that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) - C_{1,t}(\mathbf{X}^{(t)}) = 0$, almost surely. ■

The last step of the proof for Proposition 2 is inspired by [80]. Based on Lemma 4, it will be shown that the sequence $\{\nabla\check{C}_{1,t}(\mathbf{X}^{(t)}) - \nabla C_{1,t}(\mathbf{X}^{(t)})\}_{t=1}^{\infty}$ goes to zero, almost surely. Together with C_2', it follows that $\lim_{t\to\infty} C_t'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0 \; \forall \mathbf{D}$, a.s. by algorithmic construction, implying convergence of a sequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to the set of stationary points of $C(\mathbf{X})$.

By the compactness of $\mathcal{X}$, it is always possible to find a convergent subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ to a limit point $\bar{\mathbf{X}} \in \mathcal{X}$. Then, by the strong law of large numbers [28] under (a1) and equicontinuity of a family of functions $\{C_{1,t}(\cdot)\}_{t=1}^{\infty}$ due to the uniform boundedness of ∇g_1 in (p5) [14], upon restricting to the subsequence, one can have

$$\lim_{t\to\infty} C_{1,t}(\mathbf{X}^{(t)}) = \mathbb{E}_{\boldsymbol{\xi}}[g_1(\bar{\mathbf{X}}, \boldsymbol{\xi})] =: C_1(\mathbf{X}). \tag{A.29}$$

Similarly, a family of functions $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is equicontinuous due to the uniform boundedness of $\nabla\check{g}_1$ in (p5). Furthermore, $\{\check{C}_{1,t}(\cdot)\}_{t=1}^{\infty}$ is pointwisely bounded by (a1)–(a3). Thus, Arzelá-Ascoli theorem (see [14, Cor. 2.5] and [22]) implies that there exists a uniformly continuous function $\check{C}_1(\mathbf{X})$ such that $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) = \check{C}_1(\mathbf{X}) \; \forall \; \mathbf{X} \in \mathcal{X}$ and after restricting to the subsequence

$$\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) = \check{C}_1(\bar{\mathbf{X}}). \tag{A.30}$$

Furthermore, since $\check{g}_1(\mathbf{X}, \mathbf{X}^{(t-1)}, \boldsymbol{\xi}^{(t)}) \geq g_1(\mathbf{X}, \boldsymbol{\xi}^{(t)})$ as in (p1), it follows that

$$\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X}) \geq 0 \; \forall \mathbf{X}. \tag{A.31}$$

By letting $t \to \infty$ on (A.31) and combining Lemma 4 with (A.29) and (A.30), one deduces that

$$\check{C}_1(\bar{\mathbf{X}}) - C_1(\bar{\mathbf{X}}) = 0, \text{ a.s.} \tag{A.32}$$

meaning that $\check{C}_{1,t}(\mathbf{X}) - C_{1,t}(\mathbf{X})$ takes its minimum at $\bar{\mathbf{X}}$ and

$$\nabla \check{C}_1(\bar{\mathbf{X}}) - \nabla C_1(\bar{\mathbf{X}}) = 0, \text{ a.s.} \tag{A.33}$$

by the first-order optimality condition.

On the other hand, the fact that $\mathbf{X}^{(t)}$ minimizes $\check{C}_t(\mathbf{X})$ by algorithmic construction and g_2' exists (so does C_2'), yields

$$\check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X}) \quad \forall \mathbf{X} \in \mathcal{X} \tag{A.34}$$

and $\lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}^{(t)}) + C_2(\mathbf{X}^{(t)}) \leq \lim_{t\to\infty} \check{C}_{1,t}(\mathbf{X}) + C_2(\mathbf{X})$ $\forall \mathbf{X} \in \mathcal{X}$, which implies

$$\lim_{t\to\infty} \langle \nabla \check{C}_{1,t}(\mathbf{X}^{(t)}), \mathbf{D} \rangle + C_2'(\mathbf{X}^{(t)}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D}. \tag{A.35}$$

Using the result in (A.33), (A.35) can be re-written as $\langle \nabla C_1(\bar{\mathbf{X}}), \mathbf{D} \rangle + C_2'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0$ $\forall \mathbf{D}$, a.s. or

$$C'(\bar{\mathbf{X}}; \mathbf{D}) \geq 0 \quad \forall \mathbf{D}, \text{ a.s.} \tag{A.36}$$

Thus, the subsequence $\{\mathbf{X}^{(t)}\}_{t=1}^{\infty}$ asymptotically coincides with the set of stationary points of $C(\mathbf{X})$. ■

Appendix B

Derivations for Chapter 4

B.1 Derivation of the posterior conditional in (4.17)

Recalling that $p(\check{\mathbf{s}}_t|\boldsymbol{f},\sigma_\nu^2) \sim \mathcal{N}(\mathbf{W}_t^\top \boldsymbol{f}, \sigma_\nu^2 \mathbf{I}_t)$ and $p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f) \sim \mathcal{N}(\boldsymbol{\mu}_{f|z}, \boldsymbol{\Sigma}_{f|z})$, one can expand $p(\boldsymbol{f}|\check{\mathbf{s}}_t,\boldsymbol{z},\boldsymbol{\theta})$ in (4.17) to arrive at (cf. (4.18))

$$
\begin{aligned}
p(\boldsymbol{f}|\check{\mathbf{s}}_t,\boldsymbol{z},\boldsymbol{\theta}) &\propto p(\check{\mathbf{s}}_t|\boldsymbol{f},\sigma_\nu^2) p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f) \\
&\propto \exp\left[-\frac{1}{2\sigma_\nu^2}\|\check{\mathbf{s}}_t - \mathbf{W}_t^\top \boldsymbol{f}\|_2^2 - \frac{1}{2}\|\boldsymbol{f}-\boldsymbol{\mu}_{f|z}\|^2_{\boldsymbol{\Sigma}_{f|z}^{-1}} \right] \\
&\propto \exp\left[-\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Sigma}^{-1}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\boldsymbol{f} + \left(\frac{1}{\sigma_\nu^2}\check{\mathbf{s}}_t^\top \mathbf{W}_t^\top + \boldsymbol{\mu}_{f|z}^\top \boldsymbol{\Sigma}_{f|z}^{-1}\right)\boldsymbol{f} \right] \\
&= \exp\left[-\frac{1}{2}\boldsymbol{f}^\top \boldsymbol{\Sigma}^{-1}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\boldsymbol{f} + \check{\boldsymbol{\mu}}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}^\top \boldsymbol{\Sigma}^{-1}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\boldsymbol{f} \right] \\
&\propto \exp\left[-\frac{1}{2}\|\boldsymbol{f}-\check{\boldsymbol{\mu}}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}\|^2_{\boldsymbol{\Sigma}^{-1}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}} \right], \qquad \text{(B.1)}
\end{aligned}
$$

which shows that the proportionality of $p(\boldsymbol{f}|\check{\mathbf{s}}_t,\boldsymbol{z},\boldsymbol{\theta})$ follows $\mathcal{N}(\check{\boldsymbol{\mu}}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t}, \boldsymbol{\Sigma}_{f|z,\boldsymbol{\theta},\check{\mathbf{s}}_t})$. ■

B.2 Derivation of (P1) in (4.46)

At time slot τ, we seek $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ minimizing $H_{\tau+1}(\boldsymbol{f}|\check{\mathbf{s}}_{\tau+1},\boldsymbol{z},\boldsymbol{\theta})$ in (4.45), which amounts to solving

$$\max_{\substack{\mathbf{w}_{\tau+1}^{(n,n')}:\\ (n,n')\in\mathcal{M}_{\tau+1}}} \sum_{\boldsymbol{z}'\in\mathcal{Z}} \int_{\boldsymbol{\theta}',\check{\mathbf{s}}'_{\tau+1}} p(\check{\mathbf{s}}'_{\tau+1},\boldsymbol{z}',\boldsymbol{\theta}')h(\boldsymbol{z}',\boldsymbol{\theta}',\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)d\boldsymbol{\theta}'d\check{\mathbf{s}}'_{\tau+1}. \tag{B.2}$$

Then, one can show that

$$\begin{aligned}\int p(\check{\mathbf{s}}'_{\tau+1},\boldsymbol{z}',\boldsymbol{\theta}')d\check{\mathbf{s}}'_{\tau+1} &= \int_{\check{\mathbf{s}}'_{\tau+1}}\int_{\boldsymbol{f}'} p(\check{\mathbf{s}}'_{\tau+1},\boldsymbol{f}',\boldsymbol{z}',\boldsymbol{\theta}')d\boldsymbol{f}'d\check{\mathbf{s}}'_{\tau+1}\\ &\overset{(e1)}{=}\iint p(\check{s}'_{\tau+1}|\boldsymbol{f}',\boldsymbol{z}',\boldsymbol{\theta}')p(\check{\mathbf{s}}'_\tau|\boldsymbol{f}',\boldsymbol{z}',\boldsymbol{\theta}')p(\boldsymbol{f}',\boldsymbol{z}',\boldsymbol{\theta}')d\boldsymbol{f}'d\check{\mathbf{s}}'_{\tau+1}\\ &= \iint p(\boldsymbol{f}',\boldsymbol{z}',\boldsymbol{\theta}'|\check{\mathbf{s}}'_\tau)p(\check{\mathbf{s}}'_\tau)d\boldsymbol{f}'d\check{\mathbf{s}}'_\tau = \int p(\boldsymbol{z}',\boldsymbol{\theta}'|\check{\mathbf{s}}'_\tau)p(\check{\mathbf{s}}'_\tau)d\check{\mathbf{s}}'_\tau \end{aligned} \tag{B.3}$$

where $(e1)$ holds due to independence between $\check{s}'_{\tau+1}$ and $\check{\mathbf{s}}'_\tau$ after conditioning on $\{\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}\}$. By substituting (B.3) into (B.2) and recalling that $\check{\mathbf{s}}_\tau$ is given at time slot τ, finding $\mathbf{w}_{\tau+1}^{(n^*,n'^*)}$ boils down to solving

$$\max_{\substack{\mathbf{w}_{\tau+1}^{(n,n')}:\\ (n,n')\in\mathcal{M}_{\tau+1}}} \sum_{\boldsymbol{z}'\in\mathcal{Z}} \int_{\boldsymbol{\theta}'} p(\boldsymbol{z}',\boldsymbol{\theta}'|\check{\mathbf{s}}_\tau)h(\boldsymbol{z}',\boldsymbol{\theta}',\mathbf{w}_{\tau+1}^{(n,n')},\check{\mathbf{s}}_\tau)d\boldsymbol{\theta}', \tag{B.4}$$

which is (P1) in (4.46). ■

Appendix C

Derivations for Chapter 5

Here we derive the variational distributions in (5.16). Terms not related to a target variable will be lumped in a generic constant c. The iteration index ℓ will be omitted for simplicity.

C.1 Variational distribution of the SLF in (5.21)

Recall that the conditional posterior obeys $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu)p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)$. The first factor in (5.16), is expressed as

$$\ln q(\boldsymbol{f}|\boldsymbol{z}) = \sum_{i=1}^{N_g} \ln q(f_i|z_i) = \sum_{k=1}^{K} \sum_{i:\tilde{\mathbf{x}}_i \in \mathcal{A}_k} \ln q(f_i|z_i = k) \tag{C.1}$$

where $\ln q(f_i|z_i = k)$ can be written as

$$\begin{aligned} \ln q(f_i|z_i = k) &= \mathbb{E}_{-q(f_i|z_i=k)} \left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)\right] + c \\ &= \mathbb{E}_{-q(f_i|z_i=k)} \left[\ln p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu)\right] + \mathbb{E}_{-q(f_i|z_i=k)} \left[\ln p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)\right] + c. \end{aligned} \tag{C.2}$$

Each term on the RHS in (C.2) is thus given by

$$\mathbb{E}_{-q(f_i|z_i=k)} \left[\ln p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu)\right] \leftrightarrow \frac{\widetilde{\varphi}_\nu}{2} \sum_{\tau=1}^{t} \left[w_{\tau,i}^2 f_i^2 - 2\left(\check{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right) w_{\tau,i} f_i \right] \tag{C.3}$$

where $\bar{f}_j := \sum_{k=1}^{K} \breve{\zeta}_k(\tilde{\mathbf{x}}_j)\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_j)$, and

$$\begin{aligned}\mathbb{E}_{-q(f_i|z_i=k)}\left[\ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f)\right] &\leftrightarrow \mathbb{E}_{-q(f_i|z_i=k)}\left[\frac{\varphi_{f_k}}{2}\big(f_i^2 - 2\mu_{f_k}f_i\big)\right] \\ &= \frac{\widetilde{\varphi}_{f_k}}{2}\big(f_i^2 - 2\breve{m}_k f_i\big). \end{aligned} \tag{C.4}$$

After substituting (C.3) and (C.4) into (C.2), the pdf $q(f_i|z_i = k)$ can be expressed as

$$\begin{aligned} q(f_i|z_i=k) \propto \exp\Bigg\{ &-\frac{1}{2}\left(\widetilde{\varphi}_\nu \sum_{\tau=1}^{t} w_{\tau,i}^2 + \widetilde{\varphi}_{f_k}\right) f_i^2 \\ &+ \left[\widetilde{\varphi}_\nu \sum_{\tau=1}^{t}\left(\check{s}_\tau - \sum_{j\neq i} w_{\tau,j}\bar{f}_j\right) w_{\tau,i} + \widetilde{\varphi}_{f_k}\breve{m}_k\right] f_i\Bigg\}. \end{aligned} \tag{C.5}$$

By completing the square, one can readily verify that $q(f_i|z_i = k) = \mathcal{N}(\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i), \breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i))$ $\forall k$, where

$$\breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i) = \left(\widetilde{\varphi}_\nu \sum_{\tau=1}^{t} w_{\tau,i}^2 + \widetilde{\varphi}_{f_k}\right)^{-1} \tag{C.6}$$

$$\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i) = \breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i)\left[\widetilde{\varphi}_\nu \sum_{\tau=1}^{t}\left(\check{s}_\tau - \sum_{j\neq i} w_{\tau,j}\bar{f}_j\right) w_{\tau,i} + \widetilde{\varphi}_{f_k}\breve{m}_k\right]. \tag{C.7}$$

Upon defining $\bar{s}_\tau := \sum_{i=1}^{N_g} w_{\tau,i}\bar{f}_i$, $\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)$ in (C.7), it follows that

$$\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i) = \bar{f}_i + \breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i)\left[\big(\breve{m}_k - \bar{f}_i\big)\widetilde{\varphi}_{f_k} + \widetilde{\varphi}_\nu \sum_{\tau=1}^{t} w_{\tau,i}\big(\check{s}_\tau - \bar{s}_\tau\big)\right]. \qquad \blacksquare$$

C.2 Variational distribution of the hidden label field (5.22)

Since $q(\boldsymbol{z}) = \prod_{i=1}^{N_g} q(z_i)$ in (5.16) because z_i and z_j $\forall i \neq j$ are independent, we focus on the derivation of $q(z_i)$. By proportionality of the conditional posterior $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{z}; \beta)$ wrt $\boldsymbol{z}$, after singling out the terms that involve $q(z_i)$, we arrive at

$$\begin{aligned}\ln q(z_i) &= \mathbb{E}_{-q(z_i)}\left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)\right] + c \\ &\leftrightarrow \mathbb{E}_{-q(z_i)}\left[\ln p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)\right] + \mathbb{E}_{-q(z_i)}\left[\ln p(z_i|\boldsymbol{z}_{-i}; \beta)\right]. \end{aligned} \tag{C.8}$$

For $z_i = k$, each term on the RHS in (C.8) becomes

$$\begin{aligned}\mathbb{E}_{-q(z_i)}\left[\ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f)\right] &\leftrightarrow \frac{1}{2}\mathbb{E}_{-q(z_i)}\left[\ln \varphi_{f_k} - \varphi_{f_k}(f_i-\mu_{f_k})^2\right] \\ &= -\frac{\widetilde{\varphi}_{f_k}}{2}\left(\underbrace{\mathbb{E}_{-q(z_i)}\left[f_i^2\right]}_{=\breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i)+\breve{\mu}^2_{f_k}(\tilde{\mathbf{x}}_i)} - 2\breve{m}_k\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i) + \underbrace{\mathbb{E}_{-q(z_i)}\left[\mu^2_{f_k}\right]}_{=\breve{\sigma}^2_k+\breve{m}^2_k}\right) \\ &\quad + \frac{1}{2}\underbrace{\mathbb{E}_{-q(z_i)}\left[\ln\varphi_{f_k}\right]}_{=\psi(\breve{a}_k)+\ln\breve{b}_k} \end{aligned} \tag{C.9}$$

and

$$\begin{aligned}\mathbb{E}_{-q(z_i)}\left[\ln p(z_i=k|\boldsymbol{z}_{-i},\beta)\right] &\leftrightarrow \mathbb{E}_{-q(z_i)}\left[\beta\sum_{j\in\mathcal{N}(\tilde{\mathbf{x}}_i)}\delta(z_j-k)\right] \\ &= \beta\sum_{j\in\mathcal{N}(\tilde{\mathbf{x}}_i)}\breve{\zeta}_k(\tilde{\mathbf{x}}_j).\end{aligned} \tag{C.10}$$

All in all, the variational pdf $q(z_i = k)$ becomes

$$\begin{aligned}q(z_i=k) \propto \exp\Bigg\{ &-\frac{\widetilde{\varphi}_{f_k}}{2}\left[\breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i)+\breve{\mu}^2_{f_k}(\tilde{\mathbf{x}}_i) - 2\breve{m}_k\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)+\breve{\sigma}^2_k+\breve{m}^2_k\right] \\ &+\frac{1}{2}(\psi(\breve{a}_k)+\ln\breve{b}_k) + \sum_{j\in\mathcal{N}(\tilde{\mathbf{x}}_i)}\beta\breve{\zeta}_k(\tilde{\mathbf{x}}_j)\Bigg\}.\end{aligned} \tag{C.11}$$

which leads to the update rule of $q(z_i = k)$ in (5.22). ■

C.3 Variational distribution of the noise precision in (5.23)

As the conditional posterior $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)$ is proportional to $p(\check{\mathbf{s}}_t|\boldsymbol{f}, \varphi_\nu)p(\varphi_\nu)$ wrt φ_ν, we can write

$$\begin{aligned}\ln q(\varphi_\nu) &= \mathbb{E}_{-q(\varphi_\nu)}\left[\ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_t)\right] + c \\ &\leftrightarrow \mathbb{E}_{-q(\varphi_\nu)}\left[\ln p(\check{\mathbf{s}}_t|\boldsymbol{f},\varphi_\nu)\right] + \mathbb{E}_{-q(\varphi_\nu)}\left[\ln p(\varphi_\nu)\right]\end{aligned} \tag{C.12}$$

where

$$\begin{aligned}\mathbb{E}_{-q_{\varphi\nu}}\left[\ln p(\check{\mathbf{s}}_t|\boldsymbol{f},\varphi_\nu)\right] &\leftrightarrow \frac{t}{2}\ln\varphi_\nu - \frac{\varphi_\nu}{2}\|\check{\mathbf{s}}_t - \mathbf{W}_t^\top \boldsymbol{f}\|_2^2 \\ &= \frac{t}{2}\ln\varphi_\nu - \frac{\varphi_\nu}{2}\sum_{\tau=1}^{t}\check{s}_\tau^2 - 2\check{s}_\tau\overline{s}_\tau + \mathbb{E}_{-q(\varphi_\nu)}\left[(\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f})^2\right] \quad \text{(C.13)}\end{aligned}$$

and

$$\mathbb{E}_{-q(\varphi_\nu)}\left[\ln p(\varphi_\nu)\right] \leftrightarrow (a_\nu - 1)\ln\varphi_\nu - \frac{\varphi_\nu}{b_\nu}. \quad \text{(C.14)}$$

After substituting (C.13) and (C.14) into (C.12), we can easily see that $q(\varphi_\nu) = \mathcal{G}(\breve{a}_\nu, \breve{b}_\nu)$ with $\breve{a}_\nu := a_\nu + t/2$, and

$$\breve{b}_\nu := \left(\frac{1}{b_\nu} + \frac{1}{2}\sum_{\tau=1}^{t}\check{s}_\tau^2 - 2\check{s}_\tau\overline{s}_\tau + \mathbb{E}_{-q(\varphi_\nu)}\left[(\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f})^2\right]\right)^{-1} \quad \text{(C.15)}$$

where

$$\mathbb{E}_{-q(\varphi_\nu)}\left[(\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f})^2\right] = \mathrm{Var}\left[\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f}\right] + \left(\mathbb{E}_{-q(\varphi_\nu)}\left[\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f}\right]\right)^2 \quad \text{(C.16)}$$

$$= \sum_{i=1}^{N_g} w_{\tau,i}^2\left[\sum_{k=1}^{K}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\left(\breve{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \breve{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i)\right) - \overline{f}_i^2\right] + \overline{s}_\tau^2 \quad \text{(C.17)}$$

by the law of total variance on $\mathrm{Var}\left[\mathbf{w}_\tau^{(n,n')^\top}\boldsymbol{f}\right]$ [11, p. 401]. ■

C.4 Variational distribution of the field means in (5.24)

Since the conditional posterior $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)$ is proportional to $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{\mu}_{f_k})$ wrt $\boldsymbol{\mu}_{f_k}$, the entries of $\boldsymbol{\mu}_{f_k}$ are iid, we have

$$\begin{aligned}\ln q(\boldsymbol{\mu}_{f_k}) &= \mathbb{E}_{-q(\boldsymbol{\mu}_{f_k})}\left[\ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)\right] + c \\ &\leftrightarrow \mathbb{E}_{-q(\boldsymbol{\mu}_{f_k})}\left[\ln p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)\right] + \mathbb{E}_{-q(\boldsymbol{\mu}_{f_k})}\left[\sum_{k=1}^{K}\ln p(\mu_{f_k})\right] \quad \text{(C.18)}\end{aligned}$$

where

$$\mathbb{E}_{-q(\boldsymbol{\mu}_{f_k})}\left[\ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f)\right] \leftrightarrow \sum_{k=1}^{K}\sum_{i=1}^{N_g} \breve{\zeta}_k(\tilde{\mathbf{x}}_i)\widetilde{\varphi}_{f_k}\left(\mu_{f_k}^2 - 2\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)\mu_{f_k}\right) \tag{C.19}$$

and

$$\mathbb{E}_{-q(\boldsymbol{\mu}_{f_k})}\left[\sum_{k=1}^{K}\ln p(\mu_{f_k})\right] \leftrightarrow \sum_{k=1}^{K}\frac{1}{\sigma_k^2}(\mu_{f_k}^2 - 2\mu_{f_k}m_k). \tag{C.20}$$

Together with (C.19) and (C.20), $\ln q(\boldsymbol{\mu}_{f_k})$ becomes

$$\ln q(\boldsymbol{\mu}_{f_k}) \leftrightarrow \sum_{k=1}^{K}\left[\left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\widetilde{\varphi}_{f_k}\right)\mu_{f_k}^2 - 2\left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\widetilde{\varphi}_{f_k}\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)\right)\mu_{f_k}\right]. \tag{C.21}$$

After completing the square of the summand in (C.21), we find $q(\mu_{f_k}) = \mathcal{N}(\breve{m}_k, \breve{\sigma}_k^2)\ \forall k$ with

$$\breve{\sigma}_k^2 := \left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\widetilde{\varphi}_{f_k}\right)^{-1} \tag{C.22}$$

$$\breve{m}_k := \breve{\sigma}_k^2\left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\widetilde{\varphi}_{f_k}\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)\right) \tag{C.23}$$

by inspection since $q(\boldsymbol{\mu}_{f_k}) = \prod_{k=1}^{K} q(\mu_{f_k})$, as in (5.16). ■

C.5 Variational distribution of the field precisions in (5.25)

Similar to $q(\boldsymbol{\mu}_{f_k})$, the pdf $q(\boldsymbol{\varphi}_f)$ can be expressed as

$$\begin{aligned}\ln q(\boldsymbol{\varphi}_f) &= \mathbb{E}_{-q(\boldsymbol{\varphi}_f)}\left[\ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\check{\mathbf{s}}_t)\right] + c \\ &\leftrightarrow \mathbb{E}_{-q(\boldsymbol{\varphi}_f)}\left[\ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta}_f)\right] + \mathbb{E}_{-q(\boldsymbol{\varphi}_f)}\left[\sum_{k=1}^{K}\ln p(\varphi_{f_k})\right]\end{aligned} \tag{C.24}$$

by appealing to the proportionality of the posterior $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)p(\boldsymbol{\varphi}_f)$ w.r.t. $\boldsymbol{\varphi}_f$. Each term on the RHS in (C.24) can be thus expressed as

$$\mathbb{E}_{-q(\boldsymbol{\varphi}_f)}\left[\ln p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_f)\right] = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\Big[\ln\varphi_{f_k} - \varphi_{f_k}\mathbb{E}_{-q(z_i)}\left[(f_i - \mu_{f_k})^2\right]\Big] + c \tag{C.25}$$

where

$$\mathbb{E}_{-q(z_i)}\left[(f_i - \mu_{f_k})^2\right] = \breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i) + \breve{\mu}^2_{f_k}(\tilde{\mathbf{x}}_i) - 2\breve{m}_k\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i) + \breve{\sigma}^2_k + \breve{m}^2_k, \tag{C.26}$$

and

$$\mathbb{E}_{-q(\boldsymbol{\varphi}_f)}\left[\sum_{k=1}^{K}\ln p(\varphi_{f_k})\right] = \sum_{k=1}^{K}\left[(a_k - 1)\ln\varphi_{f_k} - \frac{\varphi_{f_k}}{b_k}\right] + c. \tag{C.27}$$

After substituting (C.25) and (C.27) into (C.24), $\boldsymbol{\varphi}_f$ can be shown to follow $q(\varphi_{f_k}) = \mathcal{G}(\breve{a}_k, \breve{b}_k)\ \forall k$ with

$$\breve{a}_k := a_k + \frac{1}{2}\sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i) \tag{C.28}$$

$$\breve{b}_k := \left[\frac{1}{b_k} + \frac{1}{2}\sum_{i=1}^{N_g}\breve{\zeta}_k(\tilde{\mathbf{x}}_i)\Big(\breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i) + \breve{\mu}^2_{f_k}(\tilde{\mathbf{x}}_i) - 2\breve{\mu}_{f_k}(\tilde{\mathbf{x}}_i)\breve{m}_k + \breve{\sigma}^2_k + \breve{m}^2_k\Big)\right]^{-1} \tag{C.29}$$

where we used that $q(\boldsymbol{\varphi}_f) = \prod_{k=1}^{K} q(\varphi_{f_k})$, as in (5.16). ■

C.6 Derivation of the cross-entropy in (5.46)

To establish the expression for $H(\boldsymbol{f}|\boldsymbol{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau)$ in (5.46), consider that at time slot $\tau + 1$. Similar to (5.45), we have

$$H(\boldsymbol{f}|\boldsymbol{z} = \boldsymbol{z}', \check{\mathbf{s}}_{\tau+1} = \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) \approx \frac{1}{2}\ln\left|\breve{\boldsymbol{\Sigma}}_{f|z', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau}\right| + \frac{N_g}{2}\Big(1 + \ln 2\pi\Big). \tag{C.30}$$

With $\boldsymbol{\Delta}_{w_{\tau+1}} := \text{diag}\left(\mathbf{w}_{\tau+1}^{(n,n')} \circ \mathbf{w}_{\tau+1}^{(n,n')}\right)$, and using the construction of $\breve{\sigma}^2_{f_k}(\tilde{\mathbf{x}}_i)$ in (C.6), we can write

$$\breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau+1};\hat{\boldsymbol{\theta}}_\tau} = \left[\breve{\boldsymbol{\Sigma}}^{-1}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau} + \widetilde{\varphi}_\nu \boldsymbol{\Delta}_{w_{\tau+1}}\right]^{-1} \tag{C.31}$$

from which we deduce that

$$\left|\breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau+1};\hat{\boldsymbol{\theta}}_\tau}\right|^{-1} = \left|\mathbf{I}_{N_g} + \widetilde{\varphi}_\nu \boldsymbol{\Delta}_{w_{\tau+1}} \breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau}\right| \left|\breve{\boldsymbol{\Sigma}}^{-1}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau}\right| \tag{C.32}$$

by using the matrix determinant identity lemma [35, Chapter 18]. Further substituting (C.32) into (C.30), leads to

$$\begin{aligned} H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}', \check{\mathbf{s}}_{\tau+1} = \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) \approx\ & H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}', \check{\mathbf{s}}_{\tau} = \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_\tau) \\ & - \frac{1}{2}\ln\left|\mathbf{I}_{N_g} + \widetilde{\varphi}_\nu \boldsymbol{\Delta}_{w_{\tau+1}} \breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau}\right|. \end{aligned} \tag{C.33}$$

It follows from the conditional entropy definition in (5.40) that

$$\begin{aligned} H(\boldsymbol{f}|\boldsymbol{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) \approx & \sum_{\boldsymbol{z}'\in\mathcal{Z}} \int p(\boldsymbol{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) \\ & \times \left(H(\boldsymbol{f}|\boldsymbol{z}=\boldsymbol{z}', \check{\mathbf{s}}_{\tau} = \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_\tau) - \frac{1}{2}\ln\left|\mathbf{I}_{N_g} + \widetilde{\varphi}_\nu \boldsymbol{\Delta}_{w_{\tau+1}} \breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau}\right|\right) d\check{\mathbf{s}}'_{\tau+1} \\ \overset{(e1)}{=} & H(\boldsymbol{f}|\boldsymbol{z}, \check{\mathbf{s}}_{\tau}; \hat{\boldsymbol{\theta}}_\tau) \\ & - \sum_{\boldsymbol{z}'\in\mathcal{Z}} \int p(\boldsymbol{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_\tau) \frac{1}{2}\ln\left|\mathbf{I}_{N_g} + \widetilde{\varphi}_\nu \boldsymbol{\Delta}_{w_{\tau+1}} \breve{\boldsymbol{\Sigma}}_{f|z',\check{\mathbf{s}}'_{\tau};\hat{\boldsymbol{\theta}}_\tau}\right| d\check{\mathbf{s}}'_{\tau}, \end{aligned} \tag{C.34}$$

where (e1) is obtained after marginalizing out $\check{s}_{\tau+1}$ from $p(\boldsymbol{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau)$ as the RHS of (C.33) is not a function of $\check{s}_{\tau+1}$. ■